Master's thesis
Geography
Human Geography

HIV EPIDEMIC AND ITS DRIVING FACTORS IN NAMIBIA – A SPATIAL APPROACH USING DEMOGRAPHIC AND HEALTH SURVEY DATA

Hanna Rekola

2018

Supervisor:
Markku Löytönen

UNIVERSITY OF HELSINKI
DEPARTMENT OF GEOSCIENCES AND GEOGRAPHY
DIVISION OF GEOGRAPHY

PL 64 (Gustaff Hällströmin katu 2)
00014 Helsingin yliopisto

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty	Laitos/Institution– Department
Faculty of Science	Department of Geosciences and Geography

Tekijä/Författare – Author

Hanna Rekola

Työn nimi / Arbetets titel – Title

HIV Epidemic And Its Driving Factors In Namibia – A Spatial Approach Using Demographic And Health Survey Data

Oppiaine /Läroämne – Subject

Human Geography

Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Master's thesis	January 2018	114 pp.

Tiivistelmä/Referat – Abstract

HIV and AIDS epidemic remains an issue burdening most severely developing countries of the Global South and especially Sub-Saharan Africa. Globally 37 million people are living with AIDS and from these more than half in SSA. In Namibia, estimated HIV prevalence is 13.8% compared to the average 7.1% in East and Southern Africa and global average 0.8%.

Until recently, representative data of HIV prevalence has not been available in developing countries. Currently HIV testing is included in Demographic and Health Surveys which are conducted in more than 90 developing countries around the world with the support of the DHS Program funded by USAID. These data also often include georeference, which enables also spatial analysis of the HIV epidemic.

In this study, HIV epidemic in Namibia has been studied using traditional methods utilised in population studies, which have been complemented with methods of spatial approach. Sub-regional estimates for HIV prevalence inside administrative regions have been modelled using Kernel density estimation. Factors driving the epidemic have been assessed with a logistic regression model that estimates individual's HIV risk.

In Namibia, HIV prevalence is highest in North-Central Namibia and Caprivi strip. However, according to sub-regional estimates for HIV prevalence there exist variation also inside these areas. In North-Central Namibia, high rates of HIV prevalence depended more on proximity to urban centre than population density alone. Presence of urban centres did not increase HIV prevalence everywhere in Namibia. Also inside urban areas, for example Windhoek, sub-regional estimates for HIV prevalence differed considerably.

In the logistic regression model, the following results were found: Women have 1.8 times higher odds for being HIV positive. 30-39-year olds had highest odds for being HIV positive. For men only, odds ratio was highest among 40-49-year olds. Higher educational attainment and higher wealth quintile decreased individual's HIV risk. Formerly married had higher risk for HIV. Certain language groups had higher odds for being HIV positive, these included especially Lozi and Oshiwambo. Increasing number of lifetime number of sexual partners increased individual's HIV risk. From regional factors, strong migration inflow and average HIV prevalence in the area increased individual's HIV risk.

The DHS data is available free of charge for academic purposes. Other data and all software used in this study are open source. These data and software were utilised partly in order to demonstrate that spatial approach on the HIV epidemic is possible even without vast resources. Especially in developing countries utilisation of open source opens up possibilities that would otherwise be limited.

The results of this study confirm to a large extent findings from other developing countries of the Global South where DHS data have been used in HIV studies. However, characteristics unique for the epidemic in Namibia were also found. The dynamics of the HIV epidemic are difficult to grasp and it is important that even though the situation is improving these complexities are challenged and the phenomenon further investigated.

Avainsanat – Nyckelord – Keywords

HIV epidemic, GIS, Demographic and Health Survey, DHS, HIV prevalence, sub-regional, logistic regression, driving factors

Säilytyspaikka – Förvaringställe – Where deposited

HELDA

Muita tietoja – Övriga uppgifter – Additional information'

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty	Laitos/Institution– Department
Matemaattis-luonnontieteellinen tiedekunta	Geotieteiden ja maantieteen laitos

Tekijä/Författare – Author

Hanna Rekola

Työn nimi / Arbetets titel – Title

HIV Epidemic And Its Driving Factors In Namibia – A Spatial Approach Using Demographic And Health Survey Data

Oppiaine /Läroämne – Subject

Kulttuurimaantiede

Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Pro Gradu -tutkielma	Tammikuu 2018	114 s.

Tiivistelmä/Referat – Abstract

HIV ja AIDS epidemia on edelleen kehittyviä maita vaivaava ongelma, joka rasittaa eniten Saharan eteläpuolista Afrikkaa. UNAIDS:in arvioiden mukaan globaalisti 37 miljoonaa sairastaa AIDSia ja arvioiden mukaan puolet näistä elää Saharan eteläpuolisessa Afrikassa. Namibiassa arvioitu HIV-esiintyvyyden on arvioitu olevan 13,8%. Itä- ja Etelä-Afrikassa vastaava osuus on 7,1% ja koko maailmassa 0,8%.

Aiemmin edustavia ja laadukkaita aineistoja HIV-esiintyvyyden arvioimiseksi ei ole ollut kehittyvissä maissa saatavilla. Nykyisin HIV testaukset on sisällytetty Demographic and Health Survey -ohjelman aineistoihin, joita kerätään yli 90 kehittyvässä maassa maailmanlaajuisesti. Aineistot sisältävät useimmiten myös georeferenssin, mikä mahdollistaa myös paikkatietoanalyysien tekemisen.

Tässä tutkimuksessa HIV-epidemiaa on tutkittu Namibiassa käyttäen perinteisiä väestöllisen tutkimuksen menetelmiä. Tutkimusta on täydennetty paikkatietomenetelmin. Hallinnollisten alueiden sisäisiä eroja on pyritty arvioimaan tuottamalla interpoloituja estimaatteja HIV-esiintyvyydestä käyttäen tiheys estimointia (Kernel density estimation). Epidemiaan vaikuttavia tekijöitä on pyritty tunnistamaan yksilön HIV-riskiä estimoivalla logistisella regressiomallilla.

Namibiassa HIV-esiintyvyys on korkein Pohjois-Namibiassa ja Caprivissa. Näiden alueiden sisällä estimoitu HIV-esiintyvyys kuitenkin vaihtelee merkittävästi. Pohjois-Namibiassa korkea HIV-esiintyvyys riippui enemmän kaupunkien läheisyydestä kuin pelkästä väentiheydestä. Kaupunkien läheisyys ei kuitenkaan lisännyt HIV-esiintyvyyttä kaikkialla Namibiassa. HIV-esiintyvyys myös vaihteli kaupunkien sisällä. Esimerkiksi Windhoekissa HIV-esiintyvyys vaihteli huomattavasti eri alueiden välillä.

Logistisen regressiomallin perusteella naiset ovat 1,8-kertaa todennäköisemmin HIV-positiivisia. 30-39-vuotiailla HIV-infektion todennäköisyys oli korkein. Pelkästään miehillä todennäköisyys oli korkein 40-49-vuotiailla. Korkeampi koulutustaso ja varallisuus madalsivat yksilön HIV-riskiä. Eronneiden ja leskien HIV-riski oli korkeampi kuin naimisissa olevien. Joillakin kieliryhmillä todennäköisyys HIV-positiivisuudelle oli korkeampi. Korkeampi seksikumppaneiden määrä elinaikana nosti yksilön HIV-riskiä. Alueellisista tekijöistä korkea muuttovoittoisuus ja keskimääräinen estimoitu HIV-esiintyvyys alueella nostivat yksilön HIV-riskiä.

DHS-aineistot ovat saatavilla maksutta tutkimuskäyttöön ohjelman verkkosivuilla. Muut tässä tutkimuksessa käytetyt aineistot ja ohjelmistot ovat maksuttomia ja perustuvat avoimelle lähdekoodille. Näiden aineistojen ja ohjelmistojen käytöllä haluttiin osoittaa, että tässä tutkimuksessa käytetyt tarkastelut on mahdollista toteuttaa myös ilman mittavia resursseja. Erityisesti kehitysmaissa avoimen lähdekoodin hyödyntäminen avaa mahdollisuuksia, jotka muutoin olisivat rajallisia.

Tämän tutkimuksen tulokset vahvistavat monia muualla kehittyvissä maissa vastaavilla aineistoilla toteutettujen tutkimusten havaintoja. Tutkimuksessa kuitenkin myös havaittiin, että Namibian epidemiaa määrittävät osittain myös tekijät, jotka eivät käyttäydy samoin muissa kehittyvissä maissa. HIV-epidemian tutkiminen on haastavaa ja on tärkeää, että sen tutkimista jatketaan haasteista huolimatta.

Avainsanat – Nyckelord – Keywords

HIV-epidemia, paikkatietomenetelmät, Demographic and Health Survey, DHS, HIV-esiintyvyys, alueelliset erot, logistinen regressio, vaikuttavat tekijät

Säilytyspaikka – Förvaringställe – Where deposited

HELDA

Muita tietoja – Övriga uppgifter – Additional information

Table of Contents

List of Abbreviations

AIC	Akaike's Information Criterion
AIDS	Acquired Immunodeficiency Syndrome
ARV	Anti Retro Viral treatment
AUC	Area under the Curve
DHS	Demographic and Health Survey
FDA	Food and Drug Administration
GIS	Geographic Information Science
HIS	Health Information System
HIV	Human Immunodeficiency Virus
LISA	Local indicators of spatial association
MOHSS	Ministry of Health and Social Services
NDHS	Namibia Demographic and Health Survey
NIP	National Institute of Pathology
NSA	Namibia Statistics Agency
NSF	The National Strategic Framework
PrEP	Pre-exposure prophylaxis
ROC	Receiver Operating Characteristic
SSA	Sub-Saharan Africa
UNAIDS	Joint United Nations Programme on HIV/AIDS
UN	United Nations
US	United States
WHO	World Health Organization

List of Figures

List of Tables

1 Introduction

HIV and AIDS epidemic is a global problem that still affects the demographic and socio-economic situation in many regions of the world to a great extent. According to UNAIDS (2017a) estimates, there are globally almost 37 million people living with AIDS. More than half, 19.4 million, of these people are living in sub-Saharan Africa (SSA) (UNAIDS 2017a). In many countries, also in SSA, the number of new infections has started to decrease after peaking in the late 1990s (UNAIDS 2017a). The coverage of HIV positive people receiving antiretroviral therapy has increased globally from 20 percent to 40 percent since 2010 and the number of AIDS deaths has decreased significantly (UNAIDS 2017a).

The United Nations (UN) made a political declaration in 2016 to end the AIDS epidemic by 2030 (UNAIDS 2016). The ultimate target set by this declaration is that by 2030 the epidemic would be ended and no longer be a severe health issue in any country. The declaration has also set multiple intermediate targets for the next five years. The UN recognises these targets as crucial in reaching the ultimate goal. UNAIDS monitors the development of the indicators. It has reported that even though situation is improving rapidly, more effort will have to be put into fighting against the epidemic if it is to be stopped (UNAIDS 2017b).

In developing countries of the Global South and especially in the SSA the situation regarding HIV and AIDS epidemic is most severe. In these countries, it has often been most difficult to monitor and manage the epidemic due to financial problems, lack of access to proper health care, and population and medical registers (Bushbeeth and Rivett 2004). Traditionally the epidemic has been monitored mainly by sentinel surveillance which, while a useful and valuable source of data, does not provide comprehensive and nationally representative results. Since 2003 HIV testing has been included in Demographic and Health Surveys (DHS) in selected countries. The DHS project collects survey data regularly in countries of the Global South (The DHS Program 2017). These new data have provided an unprecedented possibilities for the study of HIV and AIDS epidemic in the Global South. With the support of these nationally representative figures it has also been possible to make corrections to the results of the sentinel surveillance which seems to give consistently higher regional results than the DHS data (Fylkesness et al. 1998).

HIV and AIDS pandemic burdens countries in the SSA unevenly. From this observation rises also motivation to study the distribution of the epidemic inside national boundaries and administrative areas. A recent trend in studying the epidemic has been to include spatial perspective to the research. For a long time statistical monitoring of HIV and AIDS epidemic in developing countries had been difficult. During recent times economic and political stability has started to improve especially in Southern Africa. This, in turn, has given rise to options for more efficient management and more profound analysis. For example the DHS data has included georeference since 1986. Introduction of HIV modules in these data have made it possible to conduct spatial analysis on the distribution of the epidemic, even inside country boundaries (for example Larmarange et al. 2011; Barankanira et al. 2016; Schaefer et al. 2017). Smaller case studies with spatial perspective have been conducted especially in Southern Africa where the HIV and AIDS situation is most severe (Tanser et al. 2009; Aulagnier et al. 2011). It has been proposed that geoinformation systems and information monitoring systems can provide more effective tools for the care of individual patients as well as developing health care system in a larger scale (Busgeeth and Rivett 2004).

Namibia is one of the countries in SSA where the HIV and AIDS epidemic has been most severe. In Finland, there exists a long research tradition in the demographics of North-Central Namibia, also known as the former Ovamboland area. Unique data has been available in the form of parish registers from local Christian congregations originating from Finnish missionaries of the Evangelical Lutheran Church. This data has provided ways to study the historical demographic development in the area (Shemeikka 1999; Notkola et al. 2000, 2004; Shemeikka and Notkola 2005; Shemeikka et al. 2007; Siiskonen 2007, 2009).

While spatial analysis on HIV and AIDS epidemic has already been widely conducted in SSA, similar comprehensive analysis has not yet been done for Namibia. Moreover, the research tradition in Finland still lacks application of spatial methods to the study of HIV and AIDS epidemic in Namibia. It has not been possible earlier due to lack of comprehensive and representative HIV data with georeference. HIV testing has been included in the Namibia Demographic and Health Survey 2013 for the first time. By utilising this new data, which also includes georeference, it is possible to examine the spatial distribution of the epidemic inside national boundaries of Namibia and its administrative areas.

This study harnesses this new data and conduct spatial analysis that provides estimates of the dynamics of HIV epidemic in Namibia. Spatial methods allow the spatial variation to be estimated as exact figures for each location instead of regional averages. One of the objectives is to produce representations of the spatial variation of the epidemic on smaller scale of observation.

The most widely studied factors affecting the HIV epidemic globally will be examined in this study. These are used in a logistic regression model to estimate HIV risk for an individual person. The aim of this is to explore which demographic, socioeconomic and geographical factors affect an individual's HIV risk in Namibia. In other words, which mechanisms drive the HIV epidemic in current day Namibia? Similar approaches have been utilised for other countries in the SSA with DHS data (for example Messina et al. 2010; Chimoyi and Musenge 2014; Barankanira et al. 2016; Schaefer et al. 2017). Examining the spatial dynamics of these sociodemographic and geographical factors can also reveal why the epidemic is most prevalent where it is.

The first part of this study reviews the theoretical background of HIV and AIDS epidemic. This includes an overview of the basics of HIV and AIDS epidemic, including its nature as a medical condition as well as the global state and spread of the epidemic. In addition, means of measuring the extent of the epidemic not only in terms of regional HIV prevalence but also smaller scale applications of spatial methods will be assessed. Traditional methods and more recent new applications of measuring HIV prevalence will be presented. Finally, a variety of spatial applications that have been used to study the spatial distribution of the epidemic will be observed.

The most commonly studied factors correlating with individual's HIV risk will be examined in the theoretical background. Earlier observations of these factors will be taken into account when selecting the factors to be included in the logistic regression model of this study. In the final part of the theory review, information about Namibia as a study region is presented. The main focus will be on examining the situation in Namibia with regard to factors that had been recognised in earlier studies as contributors to HIV risk for an individual person elsewhere in the Global South. The aim is to recognise differences in the geographical distribution of these factors in Namibia and to make crude visual interpretation of how estimated sub-regional HIV prevalence and factors assumed to associate with it are connected geographically. The past and current situation regarding HIV and AIDS epidemic in Namibia, according to earlier study results as well as the NDHS 2013 final report will be presented.

2 Background

2.1 HIV and AIDS

HIV (Human Immunodeficiency Virus) is a retrovirus, first identified in 1983, that attacks the body's immune system by targeting the cells of the immune system. As a result the immune system is weakened and a person is exposed to other diseases, such as infections and cancer. The virus can lie dormant for many years before showing any symptoms. AIDS (Acquired Immunodeficiency Syndrome) is the most severe stage of HIV infection. AIDS is identified either by the presence of particular infections defined by WHO or by blood sample which shows a significant decrease in healthy cells of the immune system. HIV infection is incurable and especially in poorer countries AIDS deaths were common before appropriate treatment became more widely available (Webb 1997; UNAIDS 2017b).

The HIV infection is caused by direct human contact and exchange of bodily fluids in sexual encounter or through contaminated blood, for example in intravenous drug use. Another important mean of infection is from a mother to a baby during labour (Sepkowitz 2001; Gould and Woods 2003). Transmission from mother to baby can be prevented with the use of appropriate antiretroviral treatment which has become widely available in developing countries. HIV studies conducted in SSA focus mainly on transmission through sexual contact (MoHSS and ICF International 2014). The spreading of the disease in SSA is complex and highly dependent of the demographic, cultural and behavioural factors. In this study the theoretical background and analysis will be focused on transmission through sexual contact.

The spread of HIV and AIDS epidemic is usually measured with HIV prevalence which represents the proportion of HIV positive people among the surveyed population at any given time. HIV incidence on the other hand represents the number of newly infected people in a given population during a defined period of time. Statistics of HIV prevalence are more frequently available whereas incidence is more difficult to measure. In this study the HIV prevalence is used to examine the spread of the epidemic in Namibia.

HIV and AIDS epidemic is a global problem but it burdens most severely the Global South and especially SSA. According to an overview of history of HIV and AIDS in

Africa by Douglas Webb (1997) the pandemic spread from "the central African AIDS belt" to Southern Africa where the situation is currently most severe. This is supported also by a recent study which estimated with spatial methods the historical diffusion of the epidemic based on data from 1986 to 2003 obtained from the U.S. Census Bureau (Kalipeni and Zulu 2008).

According to UNAIDS (2017a), the estimate for global HIV prevalence in 2016 was 0.8 percent. In eastern and Southern Africa the same figure is 7.1 percent. The distribution for HIV prevalence in Africa is presented in figure 1. Table 1 also presents the six African countries that hold the highest HIV prevalence globally in 2016 according to UNAIDS. HIV prevalence is currently estimated to be highest in the most southern countries of the SSA. The lowest levels of HIV prevalence have been reported from the northern parts of the continent. It seems apparent that the regional distribution of the HIV and AIDS epidemic in SSA is highly uneven. For example in Swaziland, Lesotho and Botswana the HIV prevalence is over 20 percent, more than three times the average 7.1 percent for the eastern and Southern Africa (UNAIDS 2017a). Namibia is among the countries where HIV prevalence is high. The UNAIDS estimated the HIV prevalence in Namibia to be 13.8 percent in 2016.

Table 1. HIV-prevalence globally, in SSA, and selected countries in SSA in 2016 (UNAIDS 2017a).

Country/Region	**Adult (15-49) prevalence (%) [low estimate – high estimate]**
Global	0.8 [0.7 - 0.9]
East and Southern Africa	7.1 [6.6 - 7.6]
Swaziland	27.2 [24.9 - 29.1]
Lesotho	25.0 [22.7 - 26.5]
Botswana	21.9 [19.2 - 23.8]
South Africa	18.9 [16.6 - 21.0]
Namibia	13.8 [12.1 - 15.1]
Zimbabwe	13.5 [11.7 - 14.9]
Namibia	13.3 [12.2 - 14.5]
Zambia	12.4 [11.8 - 13.0]
Mozambique	12.3 [10.6 - 13.9]
Malawi	9.2 [8.6 - 9.7]

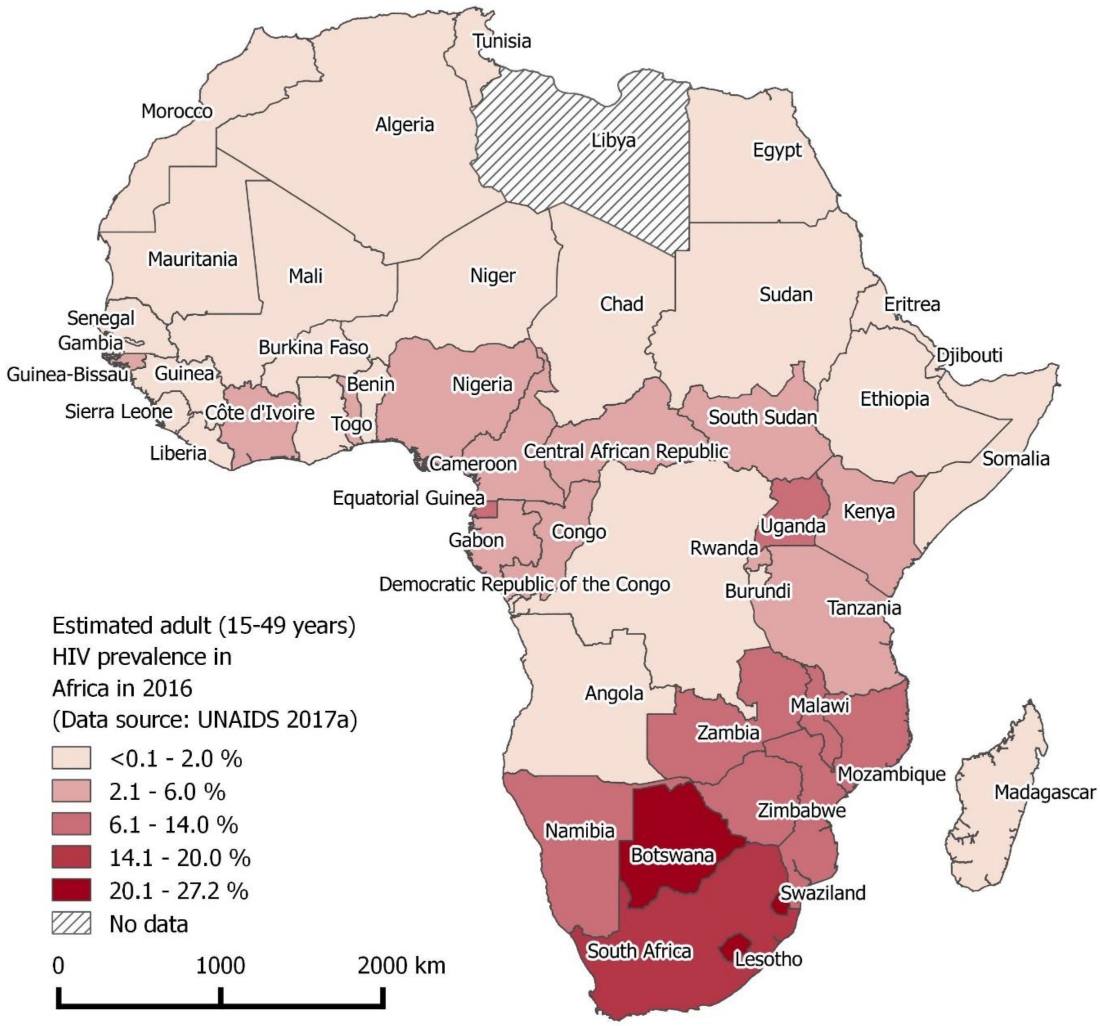

Figure 1. Estimated adult (15-49 years) HIV prevalence in Africa (Data source: UNAIDS 2017a).

The situation regarding HIV and AIDS situation has improved in many ways since the epidemic peaked in the 1990s. By the 21th century, in most countries, also in the Global South, the number of new infections and number of AIDS related deaths had dropped significantly. UNAIDS (2017b) reported that the number of new infections has decreased by 16 percent between 2010 and 2016 and AIDS related deaths by 32 percent in the same period.

In 2016, the UN made a political declaration about ending AIDS by 2030 (UNAIDS 2016). The ultimate target of the declaration is that by the year 2030 HIV and AIDS epidemic would be ended and not be a severe health issue in any country. As a part of this declaration, five-year targets were set. These targets are set for the year 2020 and they are recognised as crucial milestones in order to end the epidemic by 2030. These intermediate targets include, for example, reducing the number of new HIV infections and AIDS deaths globally to less than 500 000 and reducing new infections among children by 95 percent, eliminating HIV and AIDS stigma and ensuring that 30 million people receive treatment for HIV (UNAIDS 2016).

The clear improvements in the state of the epidemic are primarily a result of global scaling up of antiretroviral (ARV) treatment (UNAIDS 2017b). This treatment helps to maintain the immune system of the body and slows down the process of the HIV infection developing into AIDS. ARV treatment also helps to prevent the transmission of the virus from mother to child during pregnancy or labour. According to UNAIDS (2014) in countries of SSA where HIV prevalence remains highest, Namibia included, 90 percent of pregnant women living with HIV were on ARV treatment. Early ARV treatment has been found to reduce risk of HIV infection for the partner of the HIV positive individual (Cohen et al. 2011).

In 2012, the United States (US) Food and Drug Administration (FDA) approved the use of antiretroviral medication as a prevention method for HIV infection. This treatment is called pre-exposure prophylaxis (PrEP). In PrEP HIV negative people who are in high risk of infection take ARV treatment to prevent getting infected (Liu et al. 2014). The concept is the same as with preventing mother-to-child transmission. Trials have been run for this treatment for some time (Vissers et al. 2008). Currently PrEP programs are ongoing in several countries of the Global South. Others are making effort to get regulatory approvals and composing implementation plans (UNAIDS 2017c). Some have argued that the use of PrEP would give people false feeling of security and hence alter their sexual behaviour, for example increase number of sexual partners or reduce the use of condom (Vissers et al. 2008). Nevertheless, the treatment is relatively new and more research should be done to find out how this prevention method can most effectively be harnessed and also how it affects individual's risk behaviour.

The declaration to end AIDS by 2030 (UNAIDS 2016) also focuses on reaching the key populations in risk. UNAIDS recognises women, girls and young men as such groups (UNAIDS 2016). Specific targets have been set for these populations. With new treatment methods, it is even more important to recognise these risk populations. In addition to planning and targeting intervention campaigns to more effectively reach these populations, countries struggling with the effects of the epidemic could also provide these treatments for the population groups most at risk (UNAIDS 2017c).

UNAIDS is monitoring the progress that is achieved against the goals set for 2020. An update published in 2017 stated that the number of new HIV infections is declining but that more progress should be made if we wanted to end the epidemic in 2030 (UNAIDS 2017b). Indeed, the state of the HIV epidemic has improved substantially, especially in eastern and Southern Africa where the number of AIDS related deaths and new infections have dropped most (UNAIDS 2017b). Nevertheless, the epidemic is still a burden for developing countries and especially SSA.

2.2 Measuring HIV prevalence

Until recently, data on HIV prevalence in poorer countries of the Global South has been difficult to obtain. The few data that have been available have been from antenatal clinics and hospitals where pregnant women have been tested for HIV infection. These sentinel surveillance data have been the primary data source for measuring HIV prevalence in developing countries of the Global South, especially in the SSA. In many countries, the sentinel surveillance was carried out regularly, which made it possible to compile comprehensive data series for long periods of time (UNAIDS 2000; García-Calleja 2006).

During the last decade, population-based HIV screening has been carried out in several countries. Among these are the DHS studies that include HIV testing. Results from these studies have confirmed that the estimates provided by sentinel surveillance are relatively representative on national level among the general population of reproductive age (UNAIDS 2000). These new national screenings have made it possible to correct the sentinel surveillance results, which have been found to give consistently lower results than population-based HIV screenings. However, there are multiple problems and sources of bias in representativeness that have to be taken into account when sentinel surveillance data is used to estimate HIV prevalence.

The sentinel surveillance only includes testing of pregnant women in antenatal clinics, which naturally leaves out men and those women who do not attend antenatal care. There seems to be a connection between fertility and risk for HIV positivity. For example, the knowledge of one’s HIV infection could have an impact on her fertility choices. Studies have shown that HIV infection decreases fertility (Shemeikka 2005). Pregnant women have been argued to be at a greater risk of new HIV infection (Montana et al. 2007; Messina et al. 2010). Conflictingly, earlier studies have found lower HIV prevalence among pregnant women for example in Uganda (Fylkesnes et al. 1998; Gray et al. 1998). It seems evident that there has existed demand for alternative ways of collecting more valid and reliable data on the HIV epidemic.

In 2003, a nationally representative sample of HIV testing was collected in Mali as a part of the DHS program. Since then, HIV screening modules have been added to Demographic and Health Surveys in several countries. Currently the DHS data with HIV testing are the main and often only source of reliable and representative data to estimate national HIV prevalence in developing countries of the Global South (Gould and Woods 2003; Larmarange et al. 2011). These new data provide new possibilities of studying the HIV epidemic. HIV testing is carried out for individuals of different age and gender. The HIV modules of DHS data are representative on national level and the sample has been designed in such a way that it enables comparison between countries as well as over time.

Problems in estimating the HIV prevalence and incidence still pose challenges to research. In addition to lack of resources, the social resistance plays a major role. Individuals tend not to want to have themselves tested because the knowledge of the infection makes no difference to their personal lives. This is because proper treatment is not necessarily available. Instead, knowledge of infection exposes them to become stigmatised by the community. HIV and AIDS still holds a social stigma especially in many countries of the Global South and SSA (Gould and Woods 2003). The DHS program conducts HIV testing completely anonymously. The individuals tested will not themselves get to see the result of the test. Instead, they are given information about AIDS and the nearest place to have themselves tested (MoHSS and ICF International 2014). Nevertheless, it needs to be acknowledged that in the DHS studies consent for the HIV testing needs to come from the respondent and source of non-response bias does exist (MoHSS and ICF International 2014).

2.3 Spatial approach and GIS

Tables and thematic maps of national HIV prevalence as presented earlier in this study (figure 1 and table 1) give a very rough picture of the geographical distribution of the epidemic. In reality, vast differences occur inside the national borders as well as inside administrative areas. It is clear that some areas experience higher HIV prevalence and other areas relatively low HIV prevalence compared to the national average. This has been demonstrated in many studies that have utilised variety of spatial methods to estimate variation in HIV prevalence inside national boundaries (for example Montana et al. 2007; Wand and Ramjee 2010; Cuadros et al. 2017).

According to these studies, places close to each other often resemble one another in terms of intensity of HIV prevalence regardless of whether these areas are inside the same state boundaries. In other words, two communities located on opposite sides of a country border could have more similar HIV prevalence than two communities that are located in the same country or even in the same administrative region but further apart than the first two communities are. Demand for smaller scale examinations of the spatial variation in HIV prevalence in the countries of the Global South seems to exist. During the recent years, when georeferenced data with representative HIV testing has become available, spatial approach has started to be more common in the research of the HIV epidemic.

When HIV testing was mainly conducted on the national scale as part of the sentinel surveillance, there were few advanced spatial applications. The DHS data has included georeference since 1986 and there are more than 180 georeferenced datasets available on the program's website (DHS Program 2017). However, reliable and representative HIV testing samples have only been added to the DHS data recently (Larmarange et al. 2011). A large part of the HIV studies using spatial approach conducted during the last ten years have used the georeferenced DHS data which includes testing for HIV (for example Montana et al. 2007; Messina et al. 2010; Chimoyi and Musenge 2014; Barankanira et al. 2016; Cuadros et al. 2017).

Smaller scale variation in HIV prevalence can be approached for example by spatial interpolation methods (for example Montana et al. 2007; Messina et al. 2010). These methods are used to model the sub-regional variation of the HIV epidemic and to produce continuous raster maps from HIV prevalence according to known georeferenced points where HIV testing has been carried out. Interpolation can be done with various methods which all calculate predicted values for unmeasured locations based on known values in surrounding points of reference (Larmarange et al. 2011).

Larmarange (et al. 2011) has introduced a kernel density estimation method for estimating sub-regional HIV prevalence based on DHS survey data and its survey cluster points. Larmarange (2013) also provides a package for R Statistics which enables easy access to this tool for all users of HIV data from DHS. This method has been used in this study and it is introduced more profoundly later. Conducting interpolation requires comprehensive and representative data which has been hard to acquire until recently, especially in developing countries of the Global South and namely Namibia.

Spatial clusters of high and low HIV prevalence can also be recognised using a Kulldorff spatial scan statistics analysis. This very popular method of examining spatiality of the HIV epidemic was first introduced in 1997 (Kulldorff 1997). Kulldorff spatial scan test recognises areas with higher or lower HIV prevalence than the expected HIV prevalence would be without any spatial correlations. The method has been used in variety of studies with DHS HIV data (Wand and Ramjee 2010; Cuadros et al. 2013; Lakew et al. 2015).

The DHS data with HIV testing has also been used to study spatial patterns of HIV prevalence in the Democratic Republic of Congo (Messina et al. 2010). The study used spatial methods to generate interpolated surfaces of HIV prevalence and also identified hotspots of high HIV prevalence based on observed and expected values of HIV prevalence. The estimated HIV prevalence was also used to explain HIV risk for an individual person. The mean estimated HIV prevalence in a 25-kilometre buffer surrounding the location of the studied community was used as an independent variable in a logistic regression model (Messina et al. 2010). This approach is also adopted in the analysis of this study.

Factors driving the HIV epidemic will be introduced more profoundly in the following chapter. Spatial approach is also evident in studies of the connections between characteristics of individuals and societies as well as geographical factors. Georeferenced HIV data has enabled measurement of the distance from a high HIV prevalence cluster to different locations. These include, for example, primary roads, urban centres (Tanser et al. 2000; Arroyo et al. 2005, 2006) and clinics offering HIV related services (Montana et al. 2007). This last application points out clearly some of the main advantages that mapping of sub-regional variation in HIV prevalence can have.

The DHS data provides unprecedented opportunity to assess the connections between socioeconomic background information and HIV status of the survey respondents in many countries of the Global South. The socioeconomic and demographic characteristics have also been studied by accounting for the spatial nature of the epidemic. Spatial modelling of the HIV prevalence can also be done separately for different socioeconomic groups.

Factors driving the epidemic on a regional scale have been studied, for example, by using spatial applications of regression analysis. Especially in the more recent studies DHS data has often been used. DHS data is usually the most reliable nationally representative HIV data that includes georeferenced in developing countries of SSA. Spatially lagged applications of regression can be used in a linear regression model or a logistic regression. A linear regression traditionally predicts regional HIV prevalence in small administrative units, for example districts or survey enumeration areas. In these spatial applications of a regression a spatially lagged dependent variable is included on the right side of the regression equation (Zulu et al. 2014). Spatial applications of logistic regression models use spatial data to estimate individual's HIV status (Chimoyi and Musenge 2014; Niragire et al. 2015; Barankanira et al. 2016).

Besides studies using the DHS data recent case studies using spatial data from a smaller study area have been published. Number of participants in these studies can range from a couple of hundred to several thousands. For example in rural South Africa a sample of more than 12,000 participants was collected from an area covering 438 square kilometres. The homesteads of the participants were located with GPS with an accuracy of less than 2 metres (Tanser et al. 2009). A case study was also conducted in 2011 in Windhoek, Namibia where 1,753 participants were interviewed and tested for HIV infection (Aulagnier et al. 2011).

The advantages for case studies with smaller samples include better accuracy of georeference. For larger samples it is expensive to accurately locate each homestead and for the DHS studies only locations of the centroids of villages or city blocks are reported. One motivation behind less accurate georeference is protection of respondent's privacy (MoHSS and ICF International 2014). Nevertheless, smaller case studies provide little in the way of comparison between regions and countries as well as over time, unlike the nationally representative DHS data.

Spatial methods have also been applied to historical HIV surveillance data covering the whole African continent obtained from the U.S. Census Bureau (Kalipeni and Zulu 2008). The data used had been collected by the Census Bureau from various sources ranging from UNAIDS to scientific literature and even the press. Known locations and time points were geo-coded to spatial format and further interpolated to continuous map surfaces illustrating the diffusion of the epidemic from 1986 to 2003 based on the data that was collected (Kalipeni and Zulu 2008).

2.4 Factors associated with HIV prevalence

The global continental and national figures do not reveal the complexities of the HIV and AIDS epidemic. For example, national figures and maps hide behind them apparent sub-regional and intra-country variations. Georeferenced HIV data provides an opportunity to review these variations. High quality comprehensive HIV data including socioeconomic background information of the tested individuals also opens up possibilities to study the factors associated with HIV prevalence.

The HIV and AIDS epidemic is a complex phenomenon and the diffusion of the pandemic has its own dynamics that need to be taken into account. An article by Gould and Woods (2003) represents the HIV and AIDS epidemic as an exceptional infectious disease in three different dimensions. Gould and Woods (2003) distinguish a medical demographic and behavioural aspect when discussing the challenges the HIV and AIDS epidemic pose to geographic and demographic research. The infection spreads mainly through sexual encounter and also to some extent through drug use in the SSA. For this reason, the diffusion of the virus follows a unique and complex pattern that is difficult to perceive. The risk of infection for different demographic and socio-economic groups varies. In addition, the location and geographical factors have been shown to have a significant influence on HIV prevalence (Gould and Woods 2003).

When discussing the demographic dimension of the epidemic gender and age can be recognised as key factors affecting HIV risk for an individual. Because the infection spreads through sexual encounter the sexually active demographic groups have been noted to have a higher risk of infection. In some early studies, younger age groups have been shown to have higher HIV prevalence (Webb 1997). Nevertheless, a recent cross-national study of HIV prevalence and associated factors in 20 countries of SSA shows that in these areas women aged 30 to 34 are most at risk. Among men the age groups that had the highest risk were those older than 35. Teenagers aged 15 to 19 actually have relatively low risk for HIV infection. The data was collected separately in each country during 2003-2008 but the DHS data sample enables comparisons between countries and different survey rounds (Magadi and Desta 2011).

Douglas Webb (1997) wrote, already in the end of the 1990s, that the HIV prevalence rates for men and women had been noted to differ in his overview of HIV and AIDS in Africa. According to many recent studies, women tend to have higher HIV prevalence than men in SSA (Hargreaves et al. 2002; Arroyo et al. 2005; Montana et al. 2007; Messina et al. 2010; Aulagnier et al. 2011; Magadi and Desta 2011). For example, according to data collected 2007 in the Democratic Republic of Congo (Messina et al. 2010) men were 71 percent less likely to be HIV positive than women. Results also show that different sociodemographic, behavioural, and geographic factors have varying influence on the individual's risk of HIV infection depending on sex of the respondent (Messina et al. 2010; Magadi and Desta 2011). In other words, different socioeconomic groups are more prone to HIV infection depending on sex. Gould and Woods point out in their article (2003) that female prevalence rates are typically 10 to 20 percent higher for women than men in Africa.

Webb (1997) states that according to medical research women are physiologically more vulnerable to the HIV infection than men. Cultural and behavioural characteristics are also thought to have an influence on the age and gender structure of HIV positive population in SSA. For example in SSA, older men often have sexual relations with younger women which could lead to the fact that for men HIV prevalence is higher among older age groups compared to women. The age differences between partners and HIV risk have been studied recently especially in South Africa. The results were not clear, but they suggest that age-disparate relationships could increase HIV risk for younger women because they connect them to older sexual networks (Harling 2014; Maughan-Brown et al. 2014).

The spatial distribution of HIV prevalence varies also depending on sex (Montana et al. 2007; Messina et al. 2010). In other words, sub-regional estimates for HIV prevalence interpolated separately for women and men have differed significantly. This affirms that demographic factors such as sex or age should be considered when examining the geographical distribution of HIV prevalence. When studying the spatial differences in HIV prevalence it is important to take into account the demographic dimension of the epidemic. This includes examining the geographical differences in sex and age structure in different areas.

In many studies, urban centres have been found to have higher HIV prevalence than rural areas in the SSA (Buvé et al. 2002; Bärnighausen et al. 2007; Magadi and Desta 2011; Niragire et al. 2015). It has also been discovered that the discrepancies between men and women in some cases peak in urban settings and tend to be milder in rural areas. Moreover, medical research has noted that often in urban areas the subtype distribution of HIV tends to be richer, which refers to more complex epidemic and multiple sources of diffusion (Arroyo et al. 2005). Conflicting results have also been discovered. Not all urban areas have higher HIV prevalence than rural areas of countries in SSA today, and it seems that migration from rural to rural areas further affects the spread of HIV epidemic. For example, in a study conducted in Zimbabwe (Coffee et al. 2005), it was found that even though urban migrants had an escalated risk for HIV infection the spouses of rural to rural migrants actually had a higher risk for getting the infection than the spouses of rural to urban migrants. This indicates that currently the disease keeps spreading in the rural areas without the contribution of urban areas in Zimbabwe.

Moreover, distance to urban concentrations and especially capital cities has been studied as a factor that affects HIV risk for an individual. For example in rural Malawi, a study was conducted that noticed statistically significant negative association between distance to a major city and risk for HIV infection. This association was only significant for women (Feldacker et al. 2010). In the study, also increasing distance to a clinic offering HIV related services correlated negatively with HIV prevalence. In other words, HIV prevalence tended to be higher near to these facilities (Feldacker et al. 2010). Conflicting discoveries also exist. A study by Montana et al. (2007) noted that, in Kenya, the number of existing services in different regions did not respond to the estimated HIV prevalence in these areas. Areas with a high prevalence tended to have less facilities offering HIV related services.

Population mobility has been studied extensively as a factor behind spatial divergence in HIV prevalence. It seems to be one of the most difficult ones to observe. Webb (1997) describes the ways the infection has probably spread in the SSA: the oscillatory migration, which in this case refers to seasonal work-related migration, has been common in the post-independence Southern Africa. This frequent population mobility has most likely had a significant effect on the diffusion of the virus. In many African countries, conflicts and displacement have caused involuntary mobility and great numbers of refugees and returnees have been relocated various times (Webb 1997).

The connection between in-migration to urban areas and HIV prevalence has also been studied in 28 countries of the SSA during the time period from 1987 to 2005 (Voeten et al. 2010). The results of the analysis indicate that until 2000 the association between in-migration and high HIV-prevalence was very strong. After the most rapid stage of the epidemic settled also the connection between numbers of in-migrants to certain areas and their HIV prevalence seemed to disappear. The historical spreading of the virus has also been studied in eastern Africa where it was found that lack of connections between different areas had an effect on the spreading of different subtypes of HIV. For example, the Democratic Republic of Congo had a relatively low HIV prevalence even during the most intensive stage of the epidemic. It was noted that since there was little interaction between the main concentrations of population in the DRC different subtypes of the virus dominated different areas of the country. It was also discovered that the minimal connections between Ethiopia and the rest of Eastern Africa probably resulted in homogeneous HIV subtype distribution in these areas (Gray et al. 2009).

It seems evident that population mobility influences the diffusion of the HIV epidemic. Population mobility has affected the diffusion of the disease especially when the spreading of the virus was most rapid. It is unclear to what extent population mobility still affects the spreading of the disease only by linking together areas of low and high HIV prevalence. However, even in relatively recent studies it has been noticed that migration both from rural to urban areas and between rural areas has an effect on HIV prevalence (Coffee et al. 2005, 2007). In survey based studies respondents classified as frequent migrants were more likely to be HIV positive (Bärnighausen et al. 2007). It seems to be important when, why, where and with whom someone moved. It seems to be relevant that after the most intensive stage of the epidemic, above all migration seems to affect the sexual behaviour of individuals and thus make them more exposed to the infection (Coffee et al. 2005; Coffee et al. 2007).

Access to means of transportation and proximity to primary roads is assumed to correlate with higher population mobility and thus increased social interaction. Proximity to roads has been used in many studies to predict HIV prevalence for example in rural South Africa, Uganda and Malawi (Tanser et al. 2000; Arroyo et al. 2006; Feldacker et al. 2010). It has been noticed that the epidemic seems to be most intensive along heavily used roads and in these areas it is also genetically more complex which indicates intensive population mobility (Arroyo et al. 2006). The results were not consistent in all cases. For example, in rural Malawi, proximity to primary roads only correlated with increased HIV prevalence for women. Among men the association was not statistically significant (Feldacker et al. 2010).

Household's incomes and regional income level has been used to predict HIV risk in SSA. Generally HIV is perceived as a burden for the poorer countries. However, conflicting results have been found when studying connections between wealth and HIV risk (Hargreaves et al. 2002; Mishra et al. 2007; Fox 2010). According to a recent study (Fox 2010), a positive correlation was found when examining national HIV prevalence and GNI per capita values for various countries of the SSA. Wealthier countries in SSA tended to have higher HIV prevalence. On the other hand, a positive correlation was found between HIV prevalence and Gini coefficient which measures income distribution of a country's residents. More equal income distribution was found to be associated with higher level of HIV prevalence (Fox 2010). However, these studies concentrate on national average of HIV prevalence. In reality vast differences most likely exist inside national borders.

Individual's socioeconomic status and its associations with HIV risk have been explored in several studies (Hargreaves et al. 2002; Bärnighausen et al. 2007; Mishra et al. 2007). Cross-national study using DHS data with HIV testing from eight different countries of the SSA discovered that in all of the countries among adults in the wealthiest quintiles HIV prevalence was higher (Mishra et al. 2007). It seems evident that factors such as sexual risk taking and contraception explains much of this (Shelton et al. 2005). It has been also noted that when these factors are controlled for in statistical analysis the wealthiest quintiles still have equal or in some cases even higher risk for HIV (Mishra et al. 2007).

Educational attainment is a common factor indicating socioeconomic status of an individual in addition to income level. Traditionally higher HIV prevalence has been found in population groups with higher educational attainment especially in the early 1990s. Studies have proposed that this connection could have resulted from more liberal sexual behaviour among the more highly educated people (Deheneffe 1998). More recently, after the epidemic has started to improve, it has been noted that educated people responded better to intervention campaigns and programs (De Walque 2002) and higher educational attainment actually correlates with lower HIV prevalence (Fylkesnes et al 2001). A recent study of SSA indicates that individuals with only primary education would be currently most at risk to be HIV infected (Magadi and Desta 2011).

It seems that associations between socioeconomic status and HIV risk were manifold. For example in Kenya young women aged 15 to 24 with low socioeconomic status had an increased risk for HIV. However, women aged 25 to 49 and men aged 15 to 24 with higher socioeconomic status had an increased risk for HIV (Hargreaves et al. 2002). As stated in the article concerning the study, demographics affect the way socioeconomic factors influence HIV risk. It seems also evident that socioeconomic status itself does not directly affect HIV risk for an individual but rather influences behaviour indirectly.

In many studies, differences in sexual behaviour are often highlighted. However, the sexual behaviour and networks is an extremely complicated research topic. Commonly used indicators are for example multiple and concurrent sexual partners, age at first sexual intercourse and condom use. In Botswana, sexual networks of people living with HIV were studied and multiple and concurrent sexual partners and unprotected sex were reported among these respondents (Kalichman et al. 2007). In other studies, non-marital sexual activity has been found to correlate with increased risk for HIV (for example Johnson et al. 2017). In survey based studies, increased number of sexual partners has correlated with increased HIV risk (Kim et al. 2016). Nevertheless, some studies, for example a study conducted in rural Malawi, did not find corresponding association (Feldacker et al. 2010).

The association between age at first sexual intercourse and HIV status has been studied in SSA. Women who had their first sexual intercourse earlier were more likely to be HIV positive than those who initiated sexual intercourse later (Hossain 2014; Niragire et al. 2015). However, not in all cases similar connections can be found. Magadi and Desta

(2011) have observed weighted HIV prevalence in several countries of SSA and found no significant relationship between young age at initiation of sexual life and HIV risk. These results highlight the complexities that exist between HIV risk, sociodemographic and behavioural factors.

Male circumcision has been found to reduce the risk for HIV infection and it has been used as HIV prevention strategy in developing countries (Atashili 2006; Maffioli 2017). This has been established in medical studies which have shown that male circumcision can reduce risk for HIV infection by 60 to 70 percent among heterosexual men (Auvert et al. 2005). Also in survey studies where connections between HIV positivity and male circumcision have been examined among men who are not circumcised the HIV prevalence is usually higher (Magadi and Desta 2011). A recent study in Lesotho has found that traditional male circumcision performed as a part of traditional initiation ritual does not have the same beneficial effect as the medical procedure (Maffioli 2017). This should be taken into account when the effect of male circumcision to HIV risk is assessed.

Condom use has not been studied that much as predictive factor for HIV risk. It has been recommended to be studied only in high-risk populations or among HIV positive survey populations (Boerma and Weir 2005). It is clear that condom use is the most effective way to prevent HIV infection. Nevertheless, condom use is often more common in non-marital sexual relations (Kalichman et al. 2007) which in some cases correlate with higher risk for HIV infection (Johnson et al. 2017). Some studies have found correlation between condom use and lower HIV prevalence (Kim et al. 2007) but other studies have not found significant results (Johnson et al. 2017).

In regard to marital status, it has been noticed in survey studies that individuals who were formerly married but divorced or widowed had increased risk for HIV (for example Magadi and Desta 2011; Mmbaga 2013; Tenkorang 2014; Kim et al. 2016). Especially among widowed people, the HIV prevalence is often significantly higher (Magadi and Desta 2011). This could be affected by the fact that AIDS related deaths are still common in developing countries of the Global South and the spouses of the widowed people could have died from AIDS.

Both women and men living in polygynous unions have been found to experience increased risk for being HIV positive (Reniers and Tfaily 2012). However, the interactions behind this phenomenon are complex. Survey data based results from 20 African countries indicated that junior wives in polygynous unions had increased risk for HIV compared to women living in monogamous union (Reniers and Tfaily 2012). Nevertheless, not in all studies associations between polygynous union or partner concurrency and HIV risk have been found (Maher et al. 2011). Another study has also underlined that polygynous compared to monogamous union protects from HIV infection as long as the sexual partners in the union stay unchanged and no extra-marital sexual relations exist (Sawers and Isaac 2017).

Demographics including sexually active age, socioeconomic status, urban settings and mobile lifestyle have all been seen as indirect factors contributing to individual's sexual networks and behaviour (Hargreaves et al. 2002; Arroyo et al. 2005; Feldacker et al. 2010). Many studies have stated that these factors can be used to grasp the complexities of the diffusion of the epidemic. Nevertheless, it should be kept in mind that globalisation has led to rapid changes in societies and cultural behaviour both in urban settings and rural areas. It is important to take these assumptions critically and verify their impact in each context. New study results are needed constantly to give information of the changing situation and provide up to date knowledge for decision making in public health care and improvement strategies.

2.5 Study region

2.5.1 History and economy

Namibia gained independence in 1990. Before that, it had been a colony of Germany from 1884. The colonisation period was a relatively long and 1915 the area of current day Namibia was occupied by South Africa and it became a "Mandated Territory of the League of Nations" (Wallace 2011). After the First World War South Africa's authority over Namibia was established through the mandate system created by the Allied Powers. According to the mandate, South Africa was obligated to govern Namibia in a way that would promote the well-being of its inhabitants and society. In reality the oppression that had taken place during the colonial period continued under the government of South Africa (Wallace 2011).

South Africa applied its policies of racial segregation also in Namibia. The aim of the policies was to establish better-controlled reserves, regulate population mobility and collect information about the locations of the reserves and urban concentrations as well as size of the labour market. At the same time, the white population had doubled since South Africa came to rule and large quantities of land were allotted to the settlers (Wallace 2011). The former Ovamboland, Kavango and Caprivi regions were segregated from the rest of Namibia with a borderline controlled by the colonial army. This was done in order to regulate land ownership and ensure sufficient workforce for white farmers. Only little development occurred in these areas during the colonial era (Bauer 1998).

According to Wallace (2011) during the South African rule, frequent mobility of both white and black population throughout the country stayed common. Nevertheless, it has to be acknowledged that the mobility of the black population was restricted and controlled by the state (Wallace 2011). After Namibia obtained independence, the previously strictly regulated migration became uncontrolled and experienced a rapid increase (Bauer 1998).

After independence Namibia had the benefits of a well-functioning state structure and vast natural resources, for example diamonds, gold and uranium, but the previous apartheid politics also resulted in a highly dualistic society. World Bank classifies Namibia today as a middle-income country. Namibia's annual economic growth is around 5 percent. The distribution of income in Namibia is highly uneven. The Gini coefficient of Namibia has been estimated to be 0.7, which is one of the highest in the world (Fox 2010). In 2009, according to estimates 21 percent of the population lived with under 1.25 US dollars a day (World Bank 2009; Wallace 2011; MoHSS and ICF International 2014; World Bank 2015).

According to NDHS 2013 final report, the estimate for Gini coefficient was 0.42 in 2013 when according to Fox (2010) the respective estimate was 0.7. In the NDHS final report, values for Gini coefficient vary from 0.24 in urban areas to 0.45 in rural areas. When regional averages are examined, lowest values are found in Erongo and Khomas regions where the Gini coefficient is estimated to be 0.18 and 0.21 (figure 2). Highest values are found in Kavango and Ohangwena regions where the estimates are 0.51 and 0.47. In these areas, economic inequality is more present. In figure 2, regional variation in income level is examined according to percentage of population living in the lowest wealth quintile. This percentage is lowest in North-Central Namibia and Caprivi region. As much as half of the population live in lowest wealth quintile in Kawango and Ohangwena regions. In Oshana region, this percentage is 10.6 which is low compared to other regions in the northern areas. Nevertheless, for example in Khomas region only 0.5 percent of the population lives in the lowest wealth quintile. Differences between income level between different administrative regions but also between urban and rural areas seem undisputed. When in urban areas 40.3 percent of the population is estimated to fall in the highest wealth quintile in urban areas this percentage is only 2.2 percent (MoHSS and ICF International 2014).

In figure 3 regional socioeconomic differences are examined according to median years of educational attainment according to NDHS 2013 final report. Figures are given for women and men separately. Generally it seems that people are more educated in Khomas region and southern parts of Namibia. In Kunene region people seem to be least educated based on median years of educational attainment. In this region half of the population has more than six years of educational attainment. Respectively, in Khomas region half of the population has almost more than ten years of educational attainment.

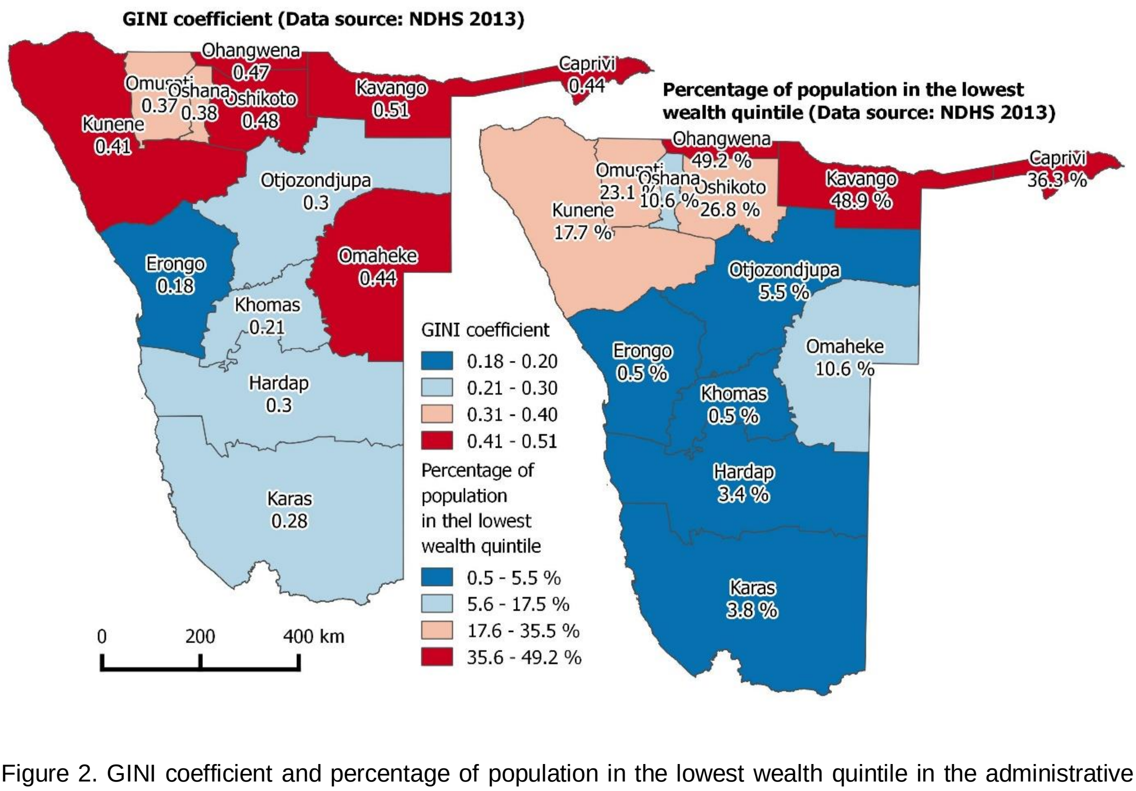

Figure 2. GINI coefficient and percentage of population in the lowest wealth quintile in the administrative regions in Namibia in 2013 (Data source: NDHS 2013).

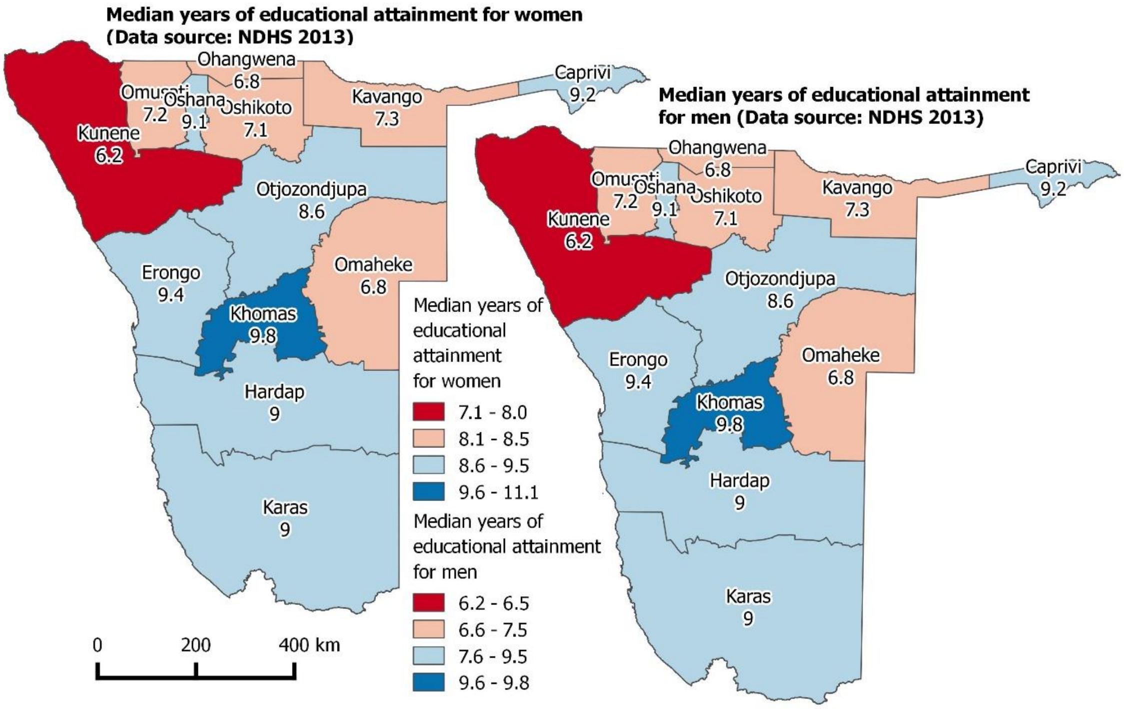

Figure 3. Median years of educational attainment for women and men in administrative regions in Namibia in 2013 (Data source: NDHS 2013).

2.5.2 *Population structure*

Namibia is the most arid country located south of the Sahara. It covers an area of approximately 824,000 square kilometres on the coast of the Atlantic Ocean. The average rainfall of 300 mm per year falls very unevenly and often as intense storms which results the infiltration to be low. The hot temperature and high altitude also cause evaporation to be high (Hellmuth 2000). Due to Namibia's ecological conditions most of the land area is arid and largely uninhabited.

The national population of Namibia according to the Population and Housing Census 2011 was 2 113 077 (NSA 2013). Population density in Namibia has been presented in figure 4 according to the census (NSA 2013). Population is highly concentrated to certain regions. These include for example North-Central Namibia, the Caprivi strip, the capital Windhoek and other larger urban concentrations (NSA 2013). The North-Central Namibia corresponds to a large part to the former Ovamboland area which consists to a large part of Omusati, Oshana, Ohangwena and Oshikoto administrative regions. This area is one of the most densely populated areas in Namibia. In less populated areas, several smaller population concentrations can be found that are limited to the existence of small towns.

The administrative areas of Namibia, the primary road network of Namibia and largest cities and towns are presented in figure 5. The location of the cities and towns are presented based on data downloaded from OpenStreetMap project (OpenStreetMap 2016). Slight modifications have been made to the data based on Namibia Population and Housing Census 2011 (NSA 2013). According to the 2011 census Windhoek is the largest city when measured by population and has a population of 325,858. The second largest city is Rundu with population of 63,431. After Rundu follow Walvis Bay and Swakopmund with populations of 63,096 and 44,725 (table 2).

When urbanisation is observed on regional and constituency scale, Khomas has the highest proportion of people living in urban areas. In Khomas region, where Windhoek is located, 95.2 percent of the population live in urban areas (NSA 2013). Khomas also has the highest total population in Namibia. The second largest region when measured by total population is Ohangwena, which has only 10.1 percent of its population living in urban areas. From the regions of North-Central Namibia only Oshana has relatively high percentage of urban population. In this region 45 percent of the total population of 176,674 live in urban areas. In Oshana is located the city of Oshakati (NSA 2013).

Table 2. Largest urban cities and towns in Namibia, their population in 2011 and population growth since 2001 (NSA 2013).

City or town	Urban population in 2011	Percentage change from 2001
Total	902 716	49.6
Windhoek	325 858	39.5
Rundu	63 431	71.6
Walvis Bay	62 096	42.4
Swakopmund	44 725	87.9
Oshakati	36 541	29.3
Katima Mulilo	28 362	28.1
Otjiwarongo	28 249	44.0
Keetmanshoop	20 977	33.0
Tsumeb	19 275	29.1
Gobabis	19 101	37.9
Mariental	12 478	26.9
Lüderitz	12 537	-5.7
Outjo	8445	40.4
Opuwo	7657	50.1
Outapi	6437	143.8
Eenhana	5528	96.4
Omuthiya	3794	-

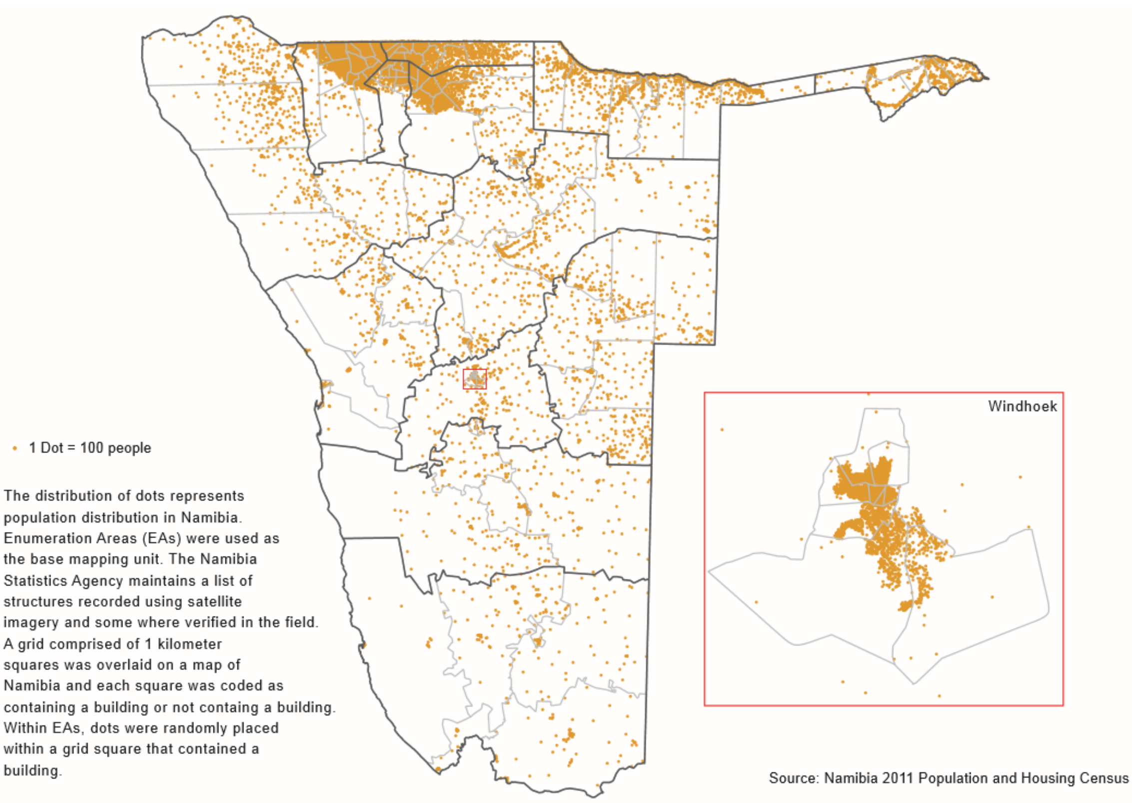

Figure 4. Population density in Namibia according to Namibia 2011 Population and Housing Census (NSA 2013).

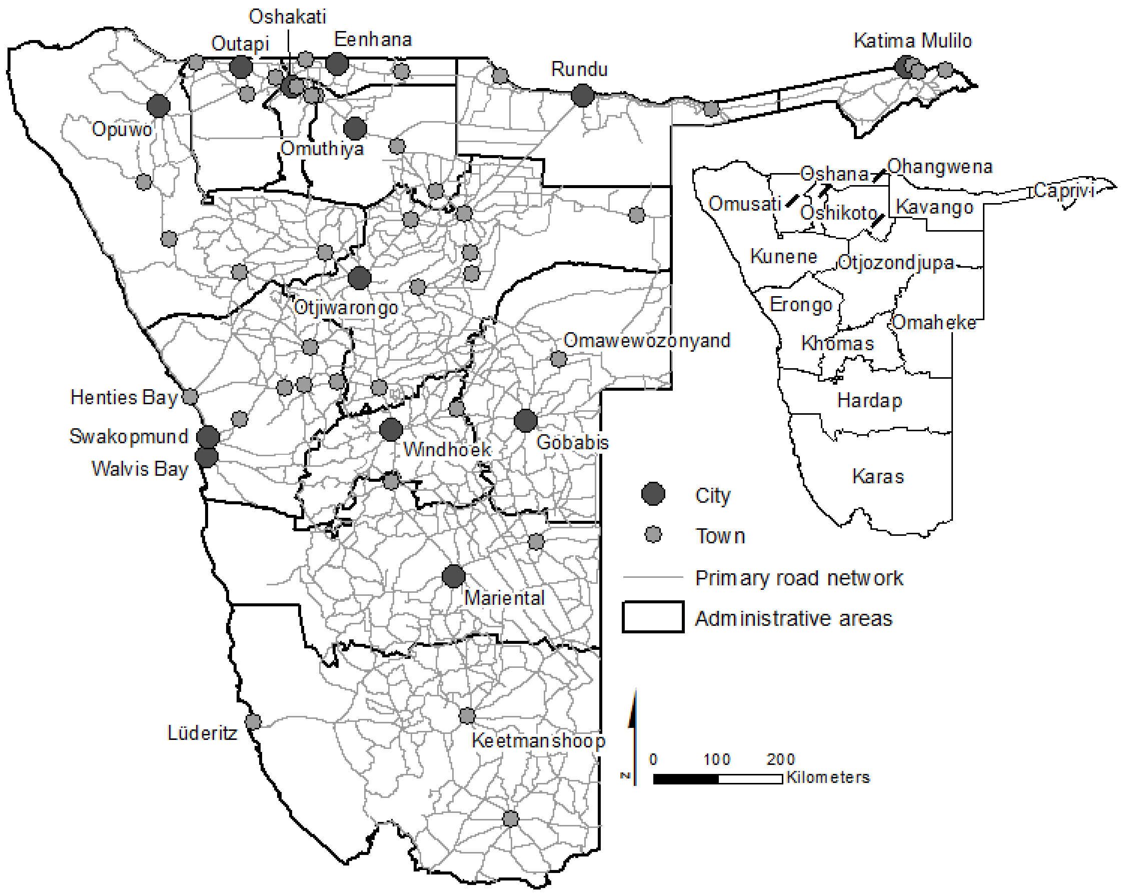

Figure 5. The primary road network of Namibia and largest cities and towns (Data source: OpenStreetMap 2016 and NSA 2013).

Regional population structure in Namibia has been reviewed here based on the Population and Housing Census 2011 (NSA 2013). In population studies, sex ratio is traditionally presented as the number of women per 100 men. The regional sex ratio in Namibia has been presented in figure 6 on a constituency level. According to the 2011 census, the northern areas of Namibia have a higher proportion of women than other areas in Namibia. Constituencies in northern Namibia can have a sex ratio as high as 130 women per 100 men whereas many constituencies in central Namibia have only around 80 to 90 women per a 100 men. Especially low proportion of women can be found for example in the surroundings of Windhoek. From the urban concentrations, Rundu forms an exception. Sex ratio for the constituencies around the city of Rundu are around 110 to 120 women per 100 men. The general sex ratio in Namibia is 106.78 according to the 2011 census (NSA 2013).

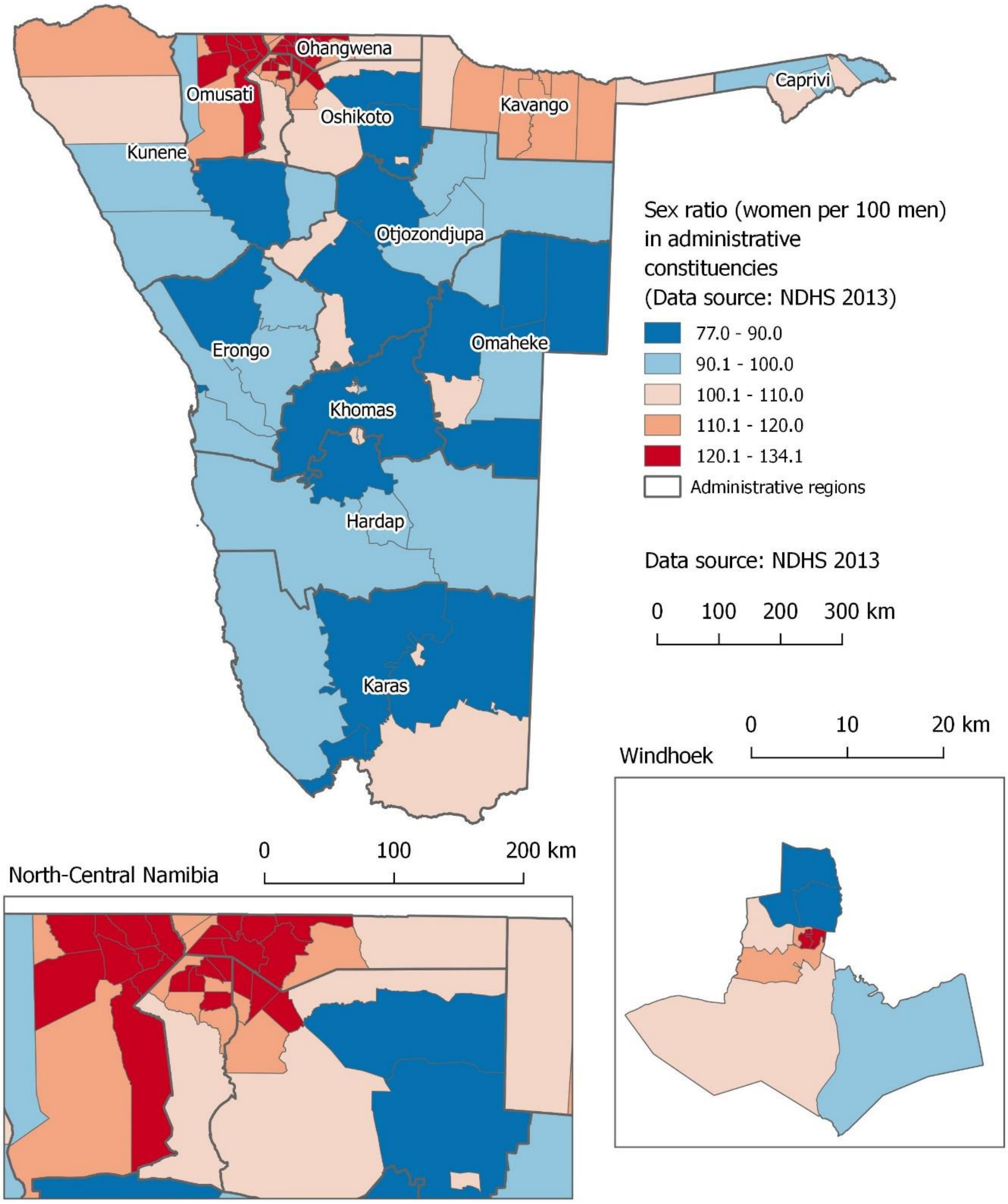

Figure 6. Sex ratio (women per 100 men) in constituencies of Namibia (Data source: NSA 2013).

According to the Population and Housing Census 2011 (NSA 2013), the median age in Namibia is 21 years. The administrative regions with highest median age are Erongo with 26 years, Khomas with 25 years and Karas with 25 years. Erongo and Khomas regions both have large urban centres. Coastal cities of Swakopmund and Walvis Bay are located in Erongo region and the capital Windhoek lies in Khomas region. In Karas region, there are no big cities but instead three smaller towns. When urban and rural locations are observed separately, a significant difference can be noticed. When urban areas have median age of 26 years, the same figure for rural areas is only 18.

In figure 7 population structure of Namibia is presented in a population pyramid based on Population and Housing Census 2011 (NSA 2013). Around 37 percent of the population in Namibia is under 15 years old. In rural areas, this percentage is 44 when in rural areas the respective percentage is 30. When men are women are observed separately, the most noticeable differences seem to occur in older age groups. For example larger proportion of people over 50 years seem to be women than men.

Regional sex and age structure have also been visualised in figures 8 to 9 as population pyramids for each of the 13 administrative areas according to the Population and Housing Census 2011 (NSA 2013). Regions with largest young age groups can be found especially in North-Central Namibia. Oshana, Oshikoto, Ohangwena and Omusati regions all have relatively large proportion of young age groups (figure 9). Also the median ages for these regions are low, around 17 to 21 years (NSA 2013). It should also be noticed that oldest age groups, ages from 80 to 90 and over, are as well larger in these regions of North-Central Namibia than respectively in other regions. In other words, working age population constitutes a relatively small part of the population in these regions compared to other regions in Namibia.

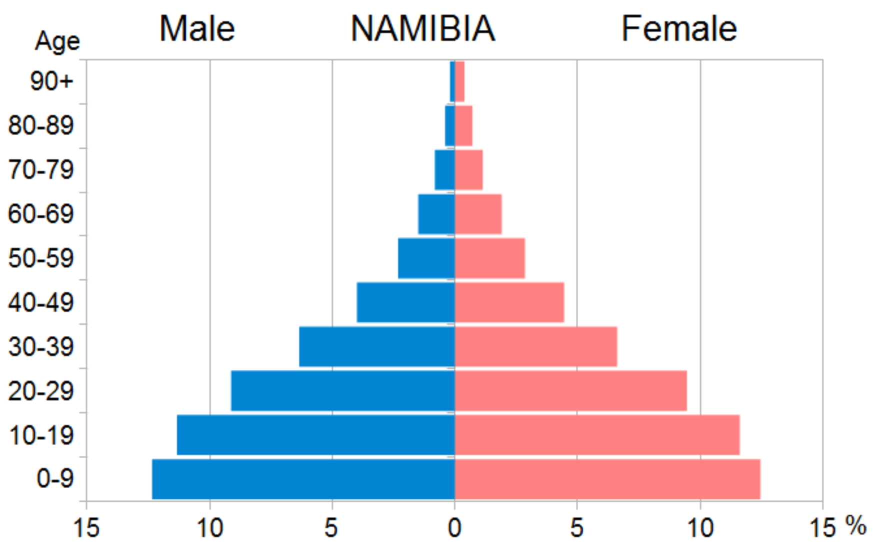

Figure 7. Population pyramid for Namibia as a whole according to Population and Housing Census 2011 (Data source: NSA 2013).

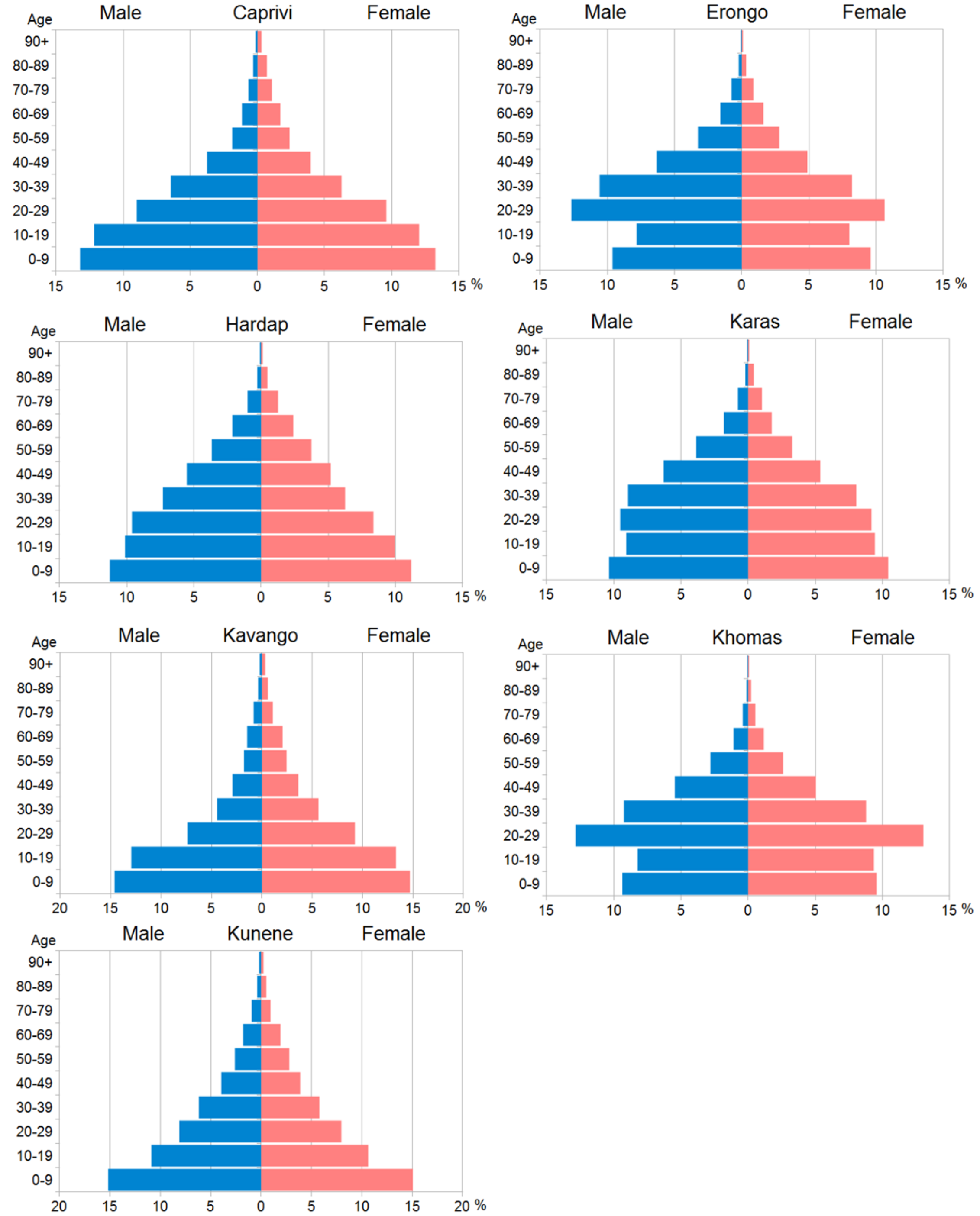

Figure 8. Population pyramids for Caprivi, Erongo, Hardap, Karas, Kavango, Khomas and Kunene administrative regions (NSA 2013).

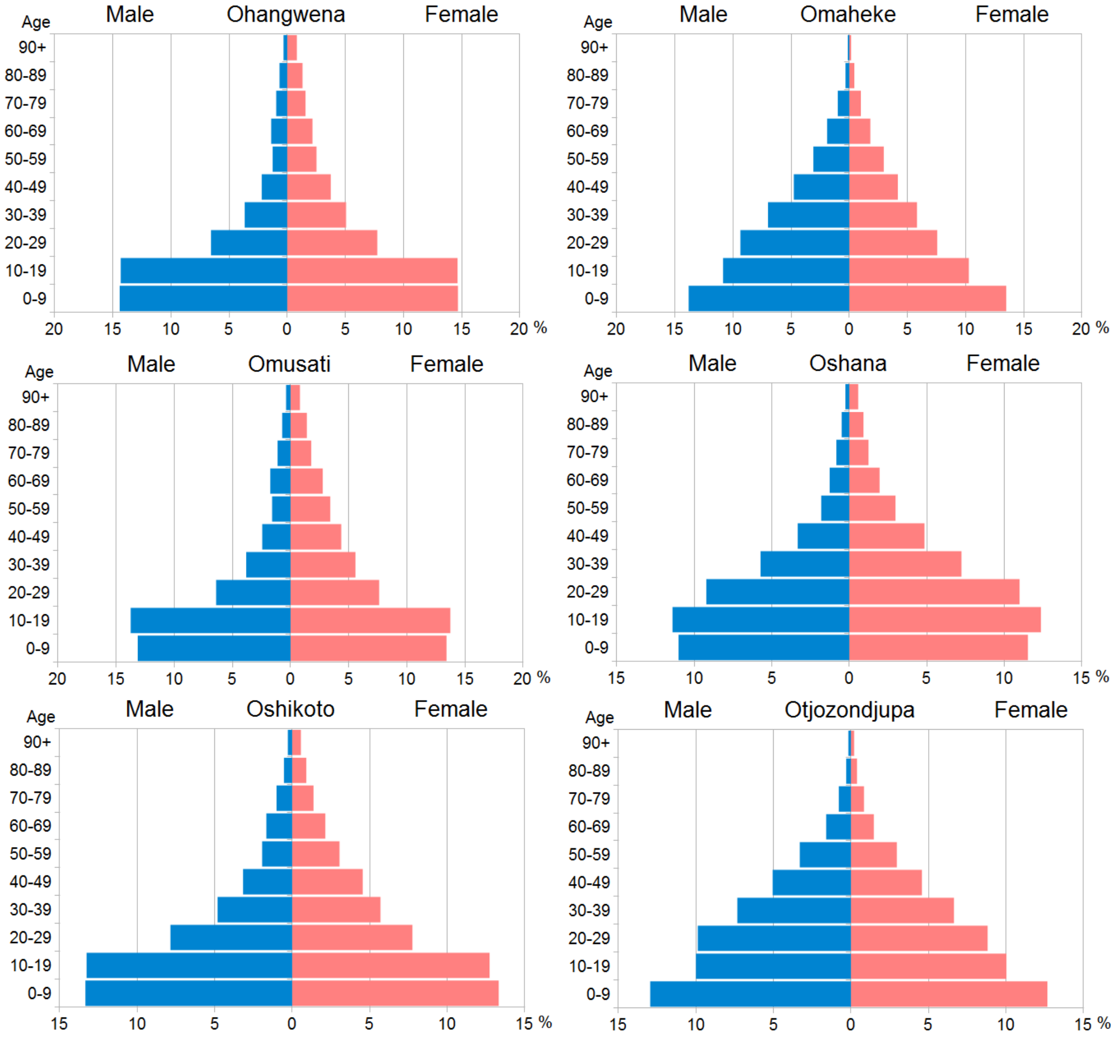

Figure 9. Population pyramids for Ohangwena, Omaheke, Omusati, Oshana, Oshikoto and Otjozondjupa administrative regions (NSA 2013).

On the other extreme are for example Erongo and Khomas regions, which have relatively high median age and also exceptionally large proportion of 20 to 29 year olds. Also Karas region has a larger proportion of 20 to 39 year olds in respect to younger age groups but the trend is not as clear as with Erongo and Khomas. It seems evident that urban areas are drawing in younger inhabitants which affects the population structure of corresponding administrative areas.

The official language in Namibia is English but people generally speak two or more languages (MoHSS and ICF International 2014). In regard to ethnic background and first spoken language, almost half of the households in Namibia spoke Oshiwambo languages. The second most common first languages are Afrikaans and Damara or Nama languages which are both spoken in around 10 percent of households in Namibia (NSA 2013).

According to the survey data of NDHS 2013, Oshiwambo languages are spoken mainly in North-Central Namibia. Afrikaans is most commonly spoken in Hardap, Karas and Erongo. Damara or Nama languages are spoken in the same regions as Afrikaans but also especially in Kunene and Omaheke. In Khomas, estimately around half of the population speaks Oshiwambo languages but Afrikaans and Damara or Nama languages are also common. Majority, little more than 40 percent, of the population is Christian and belong to the Evangelical Lutheran Church in Namibia (MoHSS and ICF International 2014).

2.5.3 Population mobility

Historically migration in Namibia was restricted and regulated by the colonial administration, first by Germany and later by South Africa. During the South African rule northern parts of Namibia, former Ovamboland, Kavango and Caprivi regions, were secluded from the rest of Namibia by colonial armies. Ethnic and racial segregation characterised the policies. Emigration was only permitted with labour contracts. Large part of the adult male population, especially of former Ovamboland region, was at one point away from their homes, working in towns, farms and mines. Population movement was highly restricted up until independence even though also before this illegal movement occurred. According to estimates, right after independence around 10 percent of people born in the former Ovamboland lived outside this region. On the opposite end, only around 2 percent born outside former Ovamboland were living in this region right after the independence (Notkola and Siiskonen 2000).

According to the Population and Housing Census 2011 (NSA 2013), a major part of the population counted in 2011 had been born in North-Central Namibia which is the most densely populated area in Namibia after Windhoek. People born in Omusati, Oshana, Ohangwena and Oshikoto regions make up 45 percent of the total population according to the census. Around 40 percent have reported their usual residence in 2011 to be in these regions, which indicates that these areas experience outflow.

Indeed according to Migration Report of Census 2011 (NSA 2015), it would appear that areas with least increase in population or even long-term decrease in respect to people born in this region can be found especially in the North-Central Namibia and around this region also in Kavango and Kunene regions (figure 10). Areas around Windhoek as well as Swakopmund and Walvis Bay by the western coast have experienced increase in population in respect to people born in those regions (figure 10). Also when short-term migration, defined by current place of residence in 2010 versus 2011, is reviewed the same areas seem to have the most population growth.

When sub-regional migration rates on constituency level are observed it can be noticed that vast differences occur for example inside North-Central Namibia. The South-Eastern part of this area has experienced most long-term increase in population compared to the number of people born in this area (figure 10). On the other hand in short-term constituencies in Oshana region have more residents in 2011 than in 2010. Inside Khomas region it seems that the northern constituencies in Windhoek have received most inflow in the long-term. On the other hand, in short-term all western constituencies in Windhoek area have received inflow.

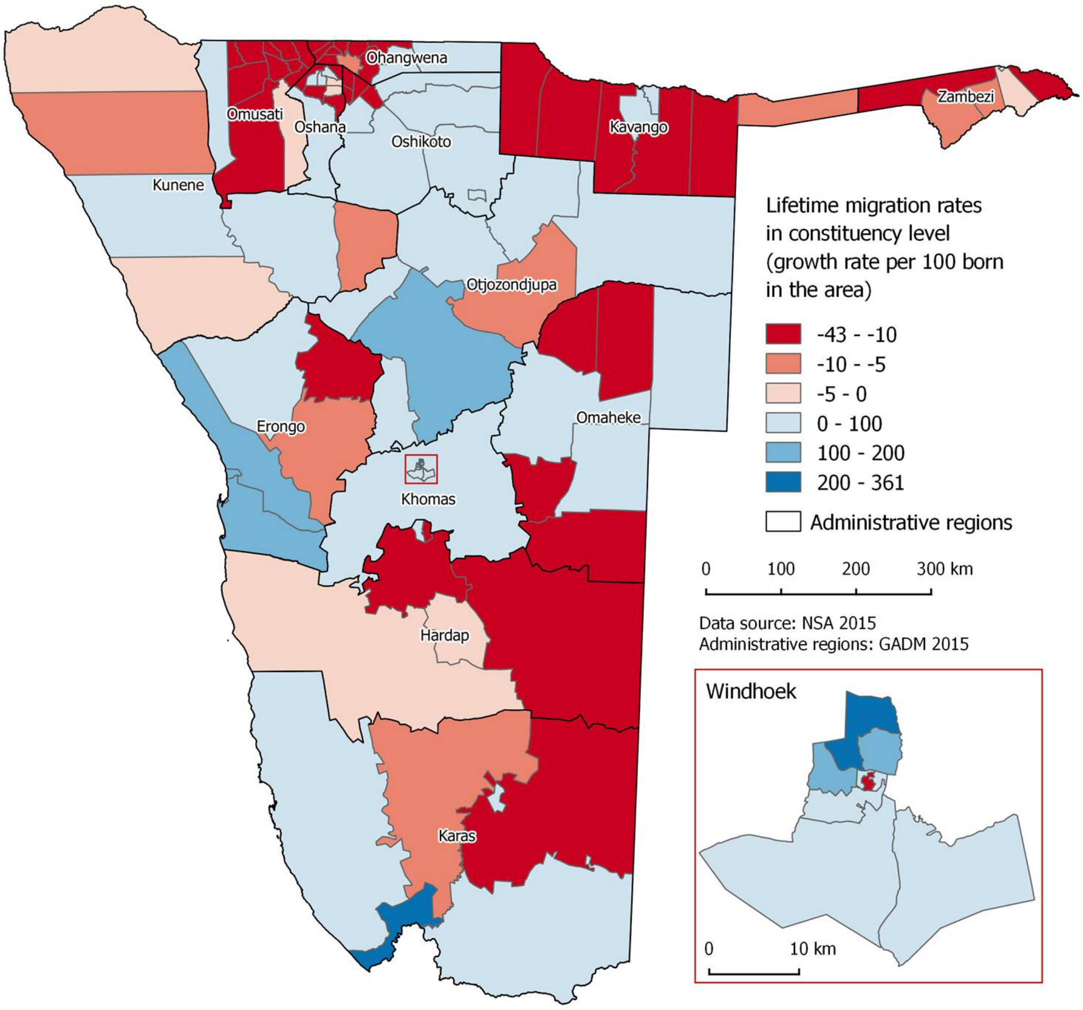

Figure 10. Lifetime migration rates in constituency level according to Migration Report of the Population and Housing Census 2011 (Data source: NSA 2015).

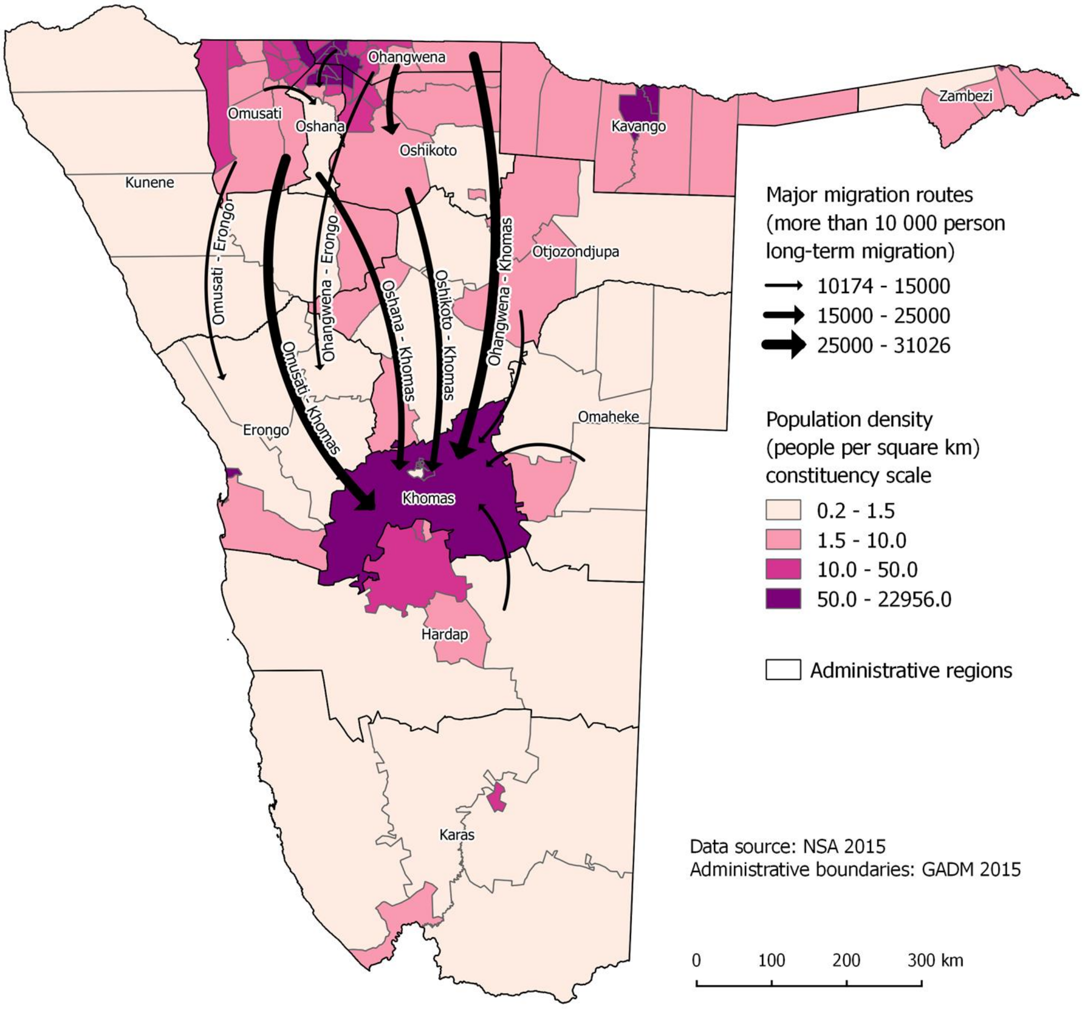

Figure 11. Major population flows defined by lifetime migration rates according to Migration Report of the Population and Housing Census 2011 (Data source: NSA 2015).

The Migration Report (NSA 2015) also specifies where from and where to people have moved on regional level. Administrative regions with least residents born in the same region are Khomas and Erongo. These regions have less than half of their current residents born in the same region. The biggest population flows between regions are presented in figure 11. The population flows are calculated indirectly as the number of people whose current region of residence and the region they were born differ. Here are presented the population flows that exceed 10,000.

The biggest population flows in the long-term occur between regions in North-Central Namibia and Khomas region (figure 11). From Ohangwena and Omusati regions, more than 25,000 people from each have moved to Khomas region. Also Erongo region has experienced inflow from North-Central Namibia. Population flows exceeding 10,000 have occurred also in-between region of North-Central Namibia for example especially from Ohangwena to Oshikoto and also from Ohangwena and Omusati to Oshana regions.

In the Population and Housing Census 2011 (NSA 2013), population growth rates have been reported for the biggest cities in Namibia. The highest growth rates are found in relatively small cities or towns where the population has nearly doubled in the last ten years. These include Eenhana and Outapi, which are located in North-Central Namibia. The biggest cities, Windhoek, Rundu and Walvis Bay, have smaller population growth rates. Windhoek and Walvis Bay experience growth rate of around 40 percent and Rundu 70 percent. When these growth rates are compared to the total population of these cities, the absolute population growth builds up to a substantial number of people moving in.

In regard to current migration patterns, also Frayne et al. (2004) have studied rural-urban social linkages and migration in Windhoek and rural Namibia through in-depth interviews and quantitative survey. According to the results of the study, majority of urban households in Windhoek have relatives in rural areas with whom they are in close contact. Urban households not only visit regularly their counterparts in rural areas but they also send their children to stay for extended times and frequently send money to relatives in rural areas. In return they receive food shipments that supplement their food budget. According to Frayne (et al. 2004) urban food security in Windhoek is relatively weak and intra-urban food sources are highly limited. Especially poor urban households are dependent on these food shipments from rural areas. These social linkages affect significantly population mobility between urban and rural areas and they are likely to exist also between other large urban centres and rural Namibia.

The Migration Report of the 2011 Census (NSA 2015) has also presented socioeconomic characteristic for a typical migrant in Namibia. According to the report, a little more than half of migrants in Namibia are young people of ages 15 to 39 years. When short-term and long-term migrations are observed separately it seems that short-term migrants seem to be on average slightly younger than life-time migrants. Men are more common migrants than women especially among lifetime migrants. In regard to marital status, migrants in Namibia tend to be never married or in consensual union. Nevertheless, in this observation age has not been standardised. Consequently it cannot be clarified whether this is only the result of younger age groups migrating more frequently (NSA 2015).

In regard to education level, people migrating tend to have higher educational attainment (NSA 2015). When employment is observed, interestingly for men employed people are more likely to migrate whereas for women the unemployed are more frequently migrants. In general, migrants tend to be employed rather than unemployed. The level of income has not been included in the Population and Housing Census 2011. Hence, the connections between unemployment and income level cannot be reviewed. Nevertheless, people with higher educational attainment could be expected to have higher income level.

2.6 HIV epidemic in Namibia and results from the Namibia Demographic and Health Survey 2013 final report

According to Webb (1997) the HIV and AIDS epidemic spread to Namibia from central Africa most probably through the Caprivi strip. Already at the turn of the 1990s the prevalence rates in Caprivi region were found to be two percentage points higher than the average in Namibia. This proposition is supported by a recent study with GIS approach which estimated historical HIV prevalence values in the African continent based on data obtained from U.S. Census Bureau (Kalipeni and Zulu 2008). According to these calculations the epidemic most likely originated from the central Africa. From there it most likely diffused southward and later to the west (Webb 1997; Kalipeni and Zulu 2008).

According to the Namibia Demographic and Health Survey (MoHSS and ICF International 2014) the HIV prevalence for the Caprivi region was 23.7 percent compared to 14.0 percent for the whole population. Also former Ovamboland area in North-Central Namibia has higher HIV prevalence compared to Namibia as a whole. Webb (1997) states that the rapid growth of the HIV epidemic in Namibia most probably originated from interaction and population mobility between Namibia and Angola and especially between Namibia and Zambia.

Siiskonen (2009) has presented the historical influence of apartheid policies and migrant workers as a contributing factor to efficient spreading of the epidemic in Southern Africa. As a result of the system men spent long periods of time away from their homes and families. Also the population mobility increased in the area and further facilitated the spreading of the virus. Also Webb (1997) proposes the cross-border movement as a crucial contributor to the rapid rise in HIV prevalence. After 1990 when Namibia gained independence more than 40,000 registered exiles returned to Namibia, 86 percent of these from Angola and 9 percent from Zambia. Webb suggests, with support of official statistics and many studies (Finckenstein 1990; Preston 1993; Simon and Preston 1993), that most probably these returnees had a significant influence on the HIV situation in Namibia. According to Webb (1997) approximately 5600 HIV infected people returned to Namibia after the independence.

It has been suggested that mortality rose in Namibia in the 1990s due to the HIV epidemic (Notkola et al. 2000). Mortality trends in Namibia after 1930 have been studied according to parish registers from missionary churches. Mortality rose in the 1990s especially among age groups where HIV prevalence has been recognised to be highest. This supports the assumption that the HIV epidemic is primarily behind the rise in mortality levels (Notkola et al. 2000, 2004).

After the rapid spread of the HIV epidemic in Namibia during the early 1990s, HIV incidence started to decline after 2003. The temporal fluctuations of the HIV prevalence among all 15 to 49 year-olds according to UNAIDS (2017a) and among pregnant women according to HIV sentinel survey (MoHSS 2014) are presented in figure 12. According to both data sources HIV prevalence grew rapidly from 1990 to 2003 when it peaked and seems to have started to decline steadily after that. It has been suggested that the sentinel surveillance data usually gives slightly higher results than national screening (for example Fylkesness et al. 1998). The results presented in figure 12 support this assumption. After 2012, the UNAIDS estimates show that HIV prevalence has increased to some extent but this trend is not showing in the sentinel surveillance data.

Recent HIV related studies in Namibia have been based on HIV sentinel surveillance data. The results of the HIV sentinel surveillance have been used in Namibia (Shemeikka 1999) where the influence of the epidemic on the demographic changes has been studied. According to the results of the survey in 1996 the proportion of HIV positive women aged 20-34 was more than 16 percent, which was the highest of all age-groups. The survey also indicated that the proportion of HIV positive women in urban areas was 17.6 percent when in rural areas the same percentage was respectively 10.4 percent. The significance of regional variation and influence of urban areas could also be seen when examining the results of the survey in smaller scale. The prevalence seemed to be highest in densely populated areas where urban centres were present (Shemeikka 1999).

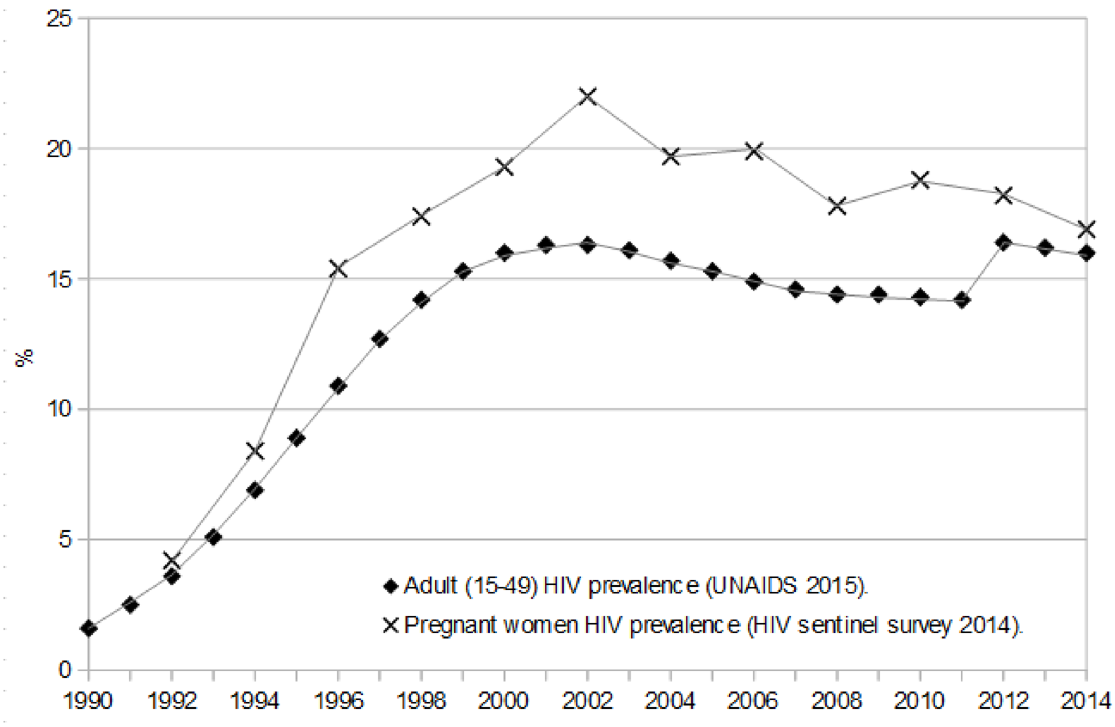

Figure 12. Adult (15-49) and pregnant women HIV prevalence in Namibia from 1990 to 2014 (Data source: UNAIDS 2017a; MoHSS 2014).

An article by Shemeikka (1999) also presents reports of the Health Information System (HIS) of the Ministry of Health and Social Services as a source of HIV data in Namibia. The system provides statistics of positive HIV tests, AIDS cases and deaths. These data support the results of the HIV sentinel surveillance data as well. Most of the HIV and AIDS cases were registered in the northern parts of Namibia and in the area surrounding Windhoek. The article recognises several problems regarding HIV and AIDS data. One problem stated is that the HIV data available lacks background variables which could help understand more profoundly the dynamics of the epidemic in Namibia. These sources of data also most probably contain bias in representativeness.

In 2013, comprehensive HIV testing was included in the Namibia Demographic and Health Survey for the first time. This new data gives an unprecedented opportunity to review the current state of the HIV epidemic in Namibia and also assess the demographic, socioeconomic and behavioural dimensions of the epidemic (MoHSS and ICF International 2014).

Namibia Demographic and Health Survey 2013 is the first nationally representative survey conducted in Namibia that includes HIV testing. For the first time, also demographic and socioeconomic characteristics of the HIV positive population can be reviewed. Key statistics and findings about the current HIV and AIDS epidemic in Namibia is available in the final report of the NDHS 2013 (MoHSS and ICF International 2014). According to the statistics in the final report, the estimated HIV prevalence in whole Namibia for the population age 15 to 49 is 14.0 percent. The UNAIDS estimate for HIV prevalence in Namibia in 2013 was 14.1 percent. In 2016 this estimate is 13.8 percent (UNAIDS 2017a).

In Namibia, the HIV prevalence for women is estimated to be significantly higher, 16.9 percent, compared to men whose prevalence rate is 10.9 percent. Notable differences between age groups were also detected (figure 13). HIV prevalence is higher for women in most age groups. Only men over 55 years have higher HIV prevalence than do women of the same age. The difference in HIV prevalence between women and men is especially high among those aged 25 to 39. Both women and men have highest HIV prevalence among 35 to 49 year olds.

The results of the NDHS 2013 final report (MoHSS and ICF International 2014) also show that sex affects the way rural versus urban place of residence influences the HIV prevalence. It seems that in Namibia for women HIV prevalence among those living in rural areas is notably higher than among those living in urban areas. With men opposite results are found even though with smaller intensity. In urban settings, men have slightly higher HIV prevalence than in rural areas. When both sexes are observed, rural areas have higher HIV prevalence, around 15.0 percent in respect to 13.3 percent in urban settings.

Regional estimates for the administrative regions in Namibia are also provided in the NDHS 2013 final report (MoHSS and ICF International 2014). The highest values for HIV prevalence are found in North-Central Namibia and the Caprivi strip (figure 14). The regions of Hardap and Omaheke in Central Namibia have the lowest HIV prevalence values. These results support earlier findings regarding geographical distribution of the epidemic in Namibia (for example Webb 1997; Shemeikka 1999).

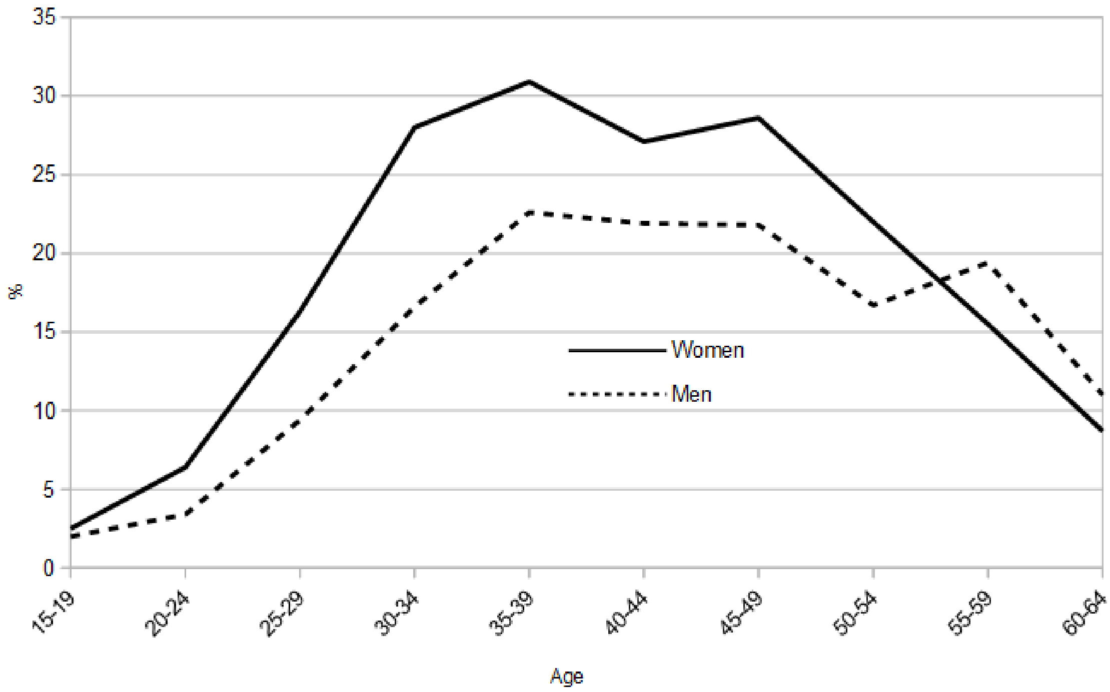

Figure 13. HIV-prevalence in Namibia according to the NDHS 2013 data among different age groups and sexes (Data source: MoHSS and ICF International 2014).

Socioeconomic characteristics of the HIV positive respondents have been presented in the NDHS 2013 final report. HIV prevalence is highest, approximately 20 percent, among groups with no educational attainment or primary education (MoHSS and ICF International 2014). Among people with secondary or higher education HIV prevalence is lower. Women seem to have higher HIV prevalence in all groups of education level excluding higher than secondary, where men have slightly higher estimate. It should be also noticed that the difference in HIV prevalence between men and women seems to be highest among groups with no educational attainment or only primary education.

In regard to household wealth quintile, women have higher HIV prevalence in all wealth quintiles. Both sexes have highest HIV prevalence in the second wealth quintile. Nevertheless, difference between second and lowest quintile seems to be larger among men where the difference is almost five percentage points and much smaller among women. Men also have higher prevalence in the middle quintile than in the lowest quintile. In other words, among men HIV risk seems to be higher if the household belongs to the second or middle wealth quintile than if it belongs to the lowest wealth quintile. For women the risk is higher for those belonging to first or second wealth quintile than for those belonging to the third quintile (MoHSS and ICF International 2014).

From those who tested positive in the HIV testing of NDHS 2013 only half reported themselves being HIV positive and currently taking ARV medication. This indicates that AIDS awareness in Namibia could be better. The National Strategic Framework (NSF) for HIV and AIDS 2010-2015 by the Ministry of Health and Social Services underlines the significance of AIDS education and awareness as a means of tackling the epidemic. It is recognised that targeting populations most at risk is especially important (MoHSS 2010).

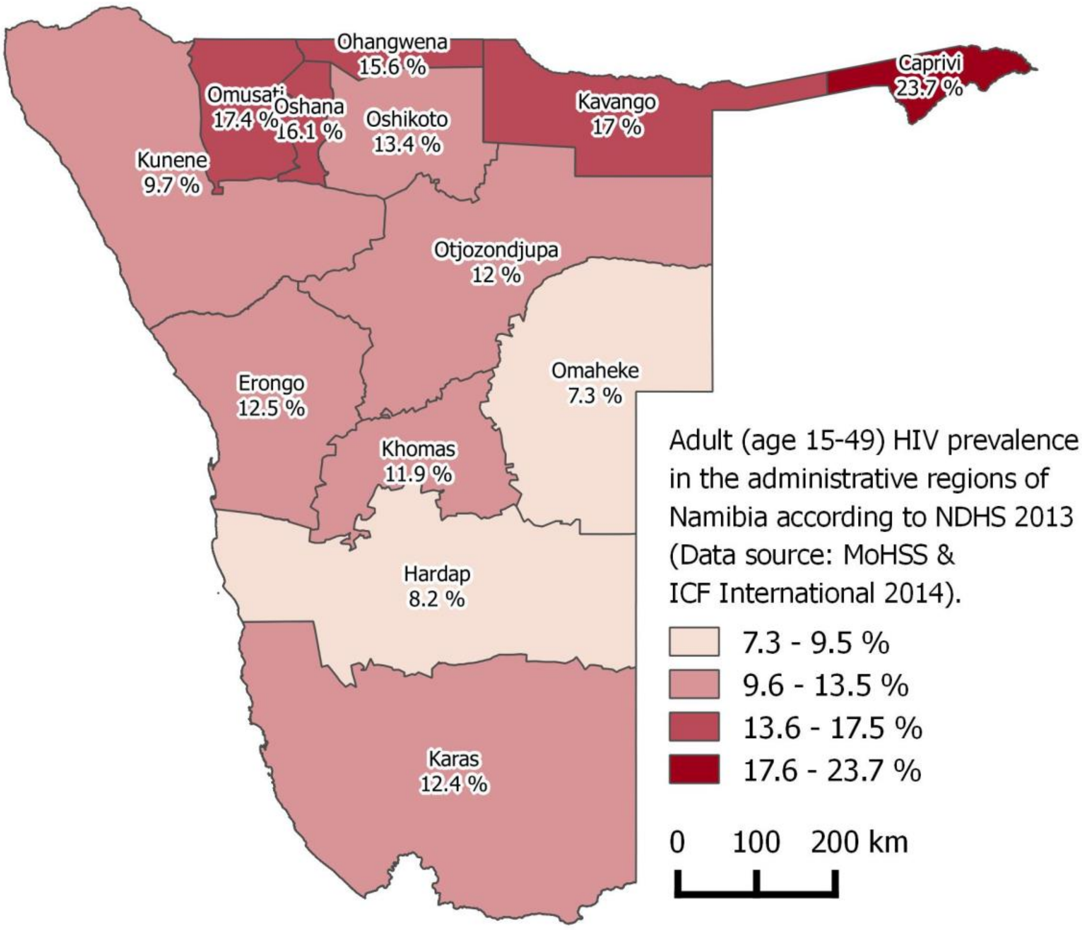

Figure 14. Regional estimates of HIV prevalence by the administrative regions of Namibia (Data source: MoHSS and ICF International 2014).

3 Objectives of the study

3.1 Key objectives

In the following part I will introduce the objectives and tentative assumptions for the analysis of this study. They include assumptions for the sub-regional spatial variation of HIV prevalence in Namibia. Assumptions have also been outlined for the logistic regression model and the selected independent factors. These assumptions have been outlined according to the theoretical background presented in previous chapters. The crude assumptions for the logistic regression model have been visualised as a chart of respective theoretical framework (figure 15).

The key objective of this study was to observe the spatial distribution of HIV prevalence in Namibia according to the recent comprehensive HIV data provided as a part of the new NDHS 2013 data released in 2015. Aim was to model sub-regional variation in the intensity of HIV epidemic in Namibia. Previous representations of the distribution of HIV epidemic in Namibia have been limited to national averages or to administrative region level figures. In this study, these boundaries have been crossed using spatial methods. Similar adaptations have been utilised earlier for different countries where georeferenced DHS data with HIV testing has been available (Montana et al. 2007; Messina et al. 2010; Larmarange et al. 2011). Similarly to preceding studies (for example Messina et al. 2010), the modelled sub-regional variations in the intensity of the HIV epidemic have also been utilised as an independent variable in a logistic regression analysis.

Another key objective was to probe what kind of factors would be currently contributing most to the HIV epidemic in Namibia. Aim was to compare selected sociodemographic and geographic factors with individual's HIV status in the NDHS survey data. It has also been examined how the influence of certain factors change when women and men are examined separately. As was shown in the previous parts of this study, many factors associate with HIV risk and the modes of interaction between the contributing factors and HIV risk are highly complex. This analysis takes after previous applications of utilising DHS data to model and explain HIV epidemic in other developing countries. Previously used methods have been used here and factors that have been noticed to have an effect on HIV epidemic in other developing countries have been included in the analysis.

In addition to producing sub-regional estimates of HIV prevalence and compiling a logistic regression model to study connections between HIV risk and contributing factors, one objective of this study was to examine the regional nature of these contributing factors and compare their spatial variation to that of the sub-regional HIV prevalence. The sociodemographic and geographical factors included in the regression model were also examined regionally in Namibia in the background part of this study. These examinations were done according to information from the NDHS 2013 and Population and Housing census 2011 final reports. These representations along with the sub-regional HIV prevalence will hopefully allow us to make crude visual interpretation of whether these factors influence the HIV epidemic the same way on regional level as they do according to the logistic regression model. It is expected that connections found in the logistic regression analysis would also be evident in spatial examinations. Nevertheless, this kind of visual interpretation can only give a very crude overall picture of the nature of the phenomenon. To acquire more reliable results, more developed spatial analysis should be conducted.

3.2 Tentative assumptions

According to earlier studies concerning HIV epidemic in Namibia (Webb 1997; Shemeikka 1999; MoHSS and ICF International 2014), HIV prevalence in Namibia is highest in the Caprivi region, Kavango region and in the North-Central Namibia, which correspond largely to the present-day Omusati, Ohangwena, Oshana and Oshikoto regions. As was pointed out in the previous chapters, historically these regions have been segregated in terms of population mobility. During the South African rule, emigration was only permitted with labour contracts and especially men worked seasonally away from home for example in farms and mines (Notkola and Siiskonen 2000). This frequent seasonal migration and limited possibilities of development in these regions have probably created idiosyncratic conditions for the spreading of the HIV and AIDS epidemic.

The assumption behind the more in-depth sub-regional spatial examinations of HIV prevalence conducted in this study is that there exists significant variation inside the borders of administrative areas. Aim is to find out whether there exists locations of especially high or low HIV prevalence inside these administrative regions that according to NDHS data (MoHSS and ICF International 2014) and other previous studies (Webb

1997; Shemeikka 1999) have high overall HIV prevalence. Sub-regional estimates for HIV prevalence are also conducted for women and men separately. This is common in HIV related studies because the dynamics of the epidemic have been noticed to differ between women and men. It is likely that the separately estimated HIV prevalence will reveal differences between women and men. It is expected that the sub-regional estimates for male HIV prevalence are lower than those for women since national HIV prevalence for men is lower. More interestingly, the examinations might reveal that hotspots of high or low HIV prevalence for men exist elsewhere than those for women.

Particular attention is paid to the northern parts of Namibia as well as large cities with high population density. According to the NDHS 2013 final report (MoHSS and ICF International 2014), in Namibia rural areas foster higher HIV prevalence than urban areas. To some extent, this is in conflict with previous study results regarding HIV epidemic in other developing counties (Bärnighausen et al. 2007; Niragire et al. 2015). The objective of this study is to study more the variation of HIV prevalence in urban and rural areas. The tentative assumption is that internal variation would be detected inside these areas. The nature of the phenomenon is most likely more complex than average values of HIV prevalence can reveal.

HIV prevalence has been demonstrated to exhibit spatial auto-correlation (for example Wand and Ramjee 2010; Cuadros et al. 2013; Lakew et al. 2015) This means that locations close to each other resemble one another in terms of HIV prevalence level. For this reason the effect of HIV prevalence in the surrounding areas was standardised in the logistic regression model. An average value of the interpolated sub-regional HIV prevalence was included in the model as a factor explaining HIV risk. In previous studies, corresponding variable has been used to estimate HIV risk for an individual in logistic regression model (Messina et al. 2010). It has been found that the surrounding overall HIV prevalence affects individual scale HIV risk. Hence, it can be assumed that similar results will also be found in this study regarding this factor.

The logistic regression analysis seeks to explain HIV risk for an individual taking into account their sociodemographic characteristic and also selected geographic factors of their surroundings. These were chosen according to previous study results where they have been found to have influence on HIV epidemic. Population density as well as population structure in respect to age and sex have been noticed to have an influence on

HIV epidemic (Webb 1997; Hargreaves et al. 2002; Arroyo et al. 2005; Montana et al. 2007; Messina et al. 2010; Aulagnier et al. 2011; Magadi and Desta 2011). According to results from previous studies, urban areas with high population density have been noticed to bear higher values of HIV prevalence. Similarly, proximity to urban centre increases HIV risk (Arroyo et al. 2005; Feldacker et al. 2010). It would be expected that similar dependencies would exist also in Namibia. However, according to the NDHS 2013 final report (MoHSS and ICF International 2014), rural areas in Namibia have higher values for HIV prevalence on survey cluster scale than urban areas. These statistics are regional averages that do not take into account variation for example between urban centre and outskirts of the city.

This being said, it still needs to be taken into account that when only men were examined the HIV prevalence was slightly higher for those residing in urban areas (MoHSS and ICF International 2014). It could be assumed that these differences would show in the individual scale logistic regression model when men and women are examined separately. Also smaller scale variation between and inside rural and urban areas can be assessed in this analysis through the modelling of sub-regional HIV prevalence. Some differences in HIV prevalence would also be expected to emerge inside areas defined as rural or urban. These differences are also most likely distinct for women and men respectively.

In respect to population structure, earlier study results have shown that HIV risk tends to be higher for women than for men in SSA (for example Arroyo et al. 2005; Montana et al. 2007; Magadi and Desta 2011). According to the NDHS 2013 final report (MoHSS and ICF International 2014), women have higher HIV prevalence in Namibia than men (figure 13). It would be expected that this would also be evident in the logistic regression model. When interpolated sub-regional HIV prevalence estimates are produced for women and men separately some differences would be expected to emerge according to regional differences in sex ratio (figure 6).

The logistic regression analysis in this study also included factors that represent individual scale population mobility and regional migration rates. According to earlier studies, population mobility and migration have been shown to have mixed effects on the HIV epidemic. At the most rapid stage of the spreading of the epidemic, population mobility had a significant effect on the patterns of diffusion. However, in present day developing countries, it seems more relevant that migration affects the behaviour, especially sexual behaviour, of individuals and thus makes them more prone to HIV risk (Coffee et al. 2005, 2007). In this analysis, both the migration status of the individual as well as regional migration rates have been examined as a part of the analysis. These two represent the phenomenon from different point of view. The logistic regression analysis also aims to probe whether just residing in an area of migration in-flow would increase HIV risk for an individual apart from them being migrants themselves.

According to the Migration Report of the Population and Housing Census (NSA 2015), most of the population movement in Namibia is directed to Khomas region and capital Windhoek. To some extent also areas around the coastal cities in the Erongo and Karas regions experience positive migration rates (figure 10). The fastest growing cities were presented in table 2. According to earlier study results, areas with migration in-flow would be expected to have high HIV prevalence in respect to other areas with lower migration rates even though earlier study results are controversial (Coffee et al. 2005, 2007; Tanser et al. 2009). Nevertheless, these regions with high migration in-flow do not bear the highest values for HIV prevalence in Namibia according to NDHS 2013 final report (MoHSS and ICF International 2014). In fact, regional HIV prevalence statistics seem to be highest in the North-Central Namibia and especially in the western parts of this area which have experienced out-flow in regard to migration rates. In this study, the impact of migration to HIV prevalence will be examined through individual scale characteristics regarding mobility and also computational migration rates. The sub-regional estimations for HIV prevalence provided in this study can also give insight into how the HIV prevalence varies within specific areas with migration in-flow.

In addition to factors representing population mobility directly also proximity to transportation network, most often primary road network, has been used in many studies to predict population mobility and thus HIV prevalence. Epidemic has been noted to be most intensive along primary roads (for example Tanser et al. 2000; Tanser et al. 2009; Arroyo et al. 2006; Feldacker et al. 2010). It needs to be taken into account that migration routes had most impact on the epidemic in the 1990s when the spreading was most intensive. It is unclear to what extent and in which ways the migration routes and intensity still affect the HIV epidemic in Namibia. One objective of this study is to probe this complex phenomenon and hopefully find new information to supplement the current knowledge on the epidemic.

In respect to individual scale socioeconomic characteristics, income level and the level of educational attainment have been examined in the background literature chapter and also included in the analysis of this study. The objective is to find out whether a person's income level or level of education affects their HIV risk in the logistic regression model. According to previous study results, individual scale income level has mixed effects on HIV risk. In some studies, it has been found that higher income level correlates with increased HIV risk (Shelton et al. 2005; Mishra et al. 2007). On the other hand, in a study conducted in Kenya, it was detected that HIV risk was especially high for young women with low socioeconomic status (Hargreaves et al. 2002). High educational attainment has traditionally been connected with increased HIV risk but more recent study results have suggested that among more educated people intervention and education in regard to HIV and AIDS have had more impact (Deheneffe 1998; De Walque 2002; Magadi and Desta 2011). It also seems evident that demographics have an effect on how the socioeconomic status affects sexual behaviour and thus HIV risk.

In Namibia, women have higher HIV prevalence in lower wealth quintiles compared to men (MoHSS and ICF International 2014). This supports the study results detected in Kenya (Hargreaves et al. 2002). When income level is examined on regional scale, highest percentage of population in the highest wealth quintiles in Namibia can be found in Khomas and Erongo regions where there are growing urban centres. The highest percentage of population in lowest wealth quintiles can be found in Ohangwena, Kawango and Caprivi regions. According to previous study results, these regions with high percentage of people in the low wealth quintiles experience higher HIV prevalence. HIV prevalence for example in Khomas region is lower but sub-regional variation can also be expected to occur (MoHSS and ICF International 2014).

According to theory background, the HIV epidemic in SSA is affected by many sociodemographic and geographical factors. It has been pointed out in many of these studies that most likely contributing factors affect HIV risk only indirectly through altered sexual behaviour. One of the key objectives of this study was to find out how significant effect individual's sexual behaviour has to their HIV risk. Factors used in this study indicating sexual behaviour include the number of sexual partners during survey respondent's lifetime and condom use.

Increasing number of sexual partners would be expected to increase individual's HIV risk (Kim et al. 2007). However, the effect of condom use to HIV risk is more difficult to interpret. Condom use physically reduces risk for HIV infection and some studies have shown that condom use correlates with lower HIV prevalence (Kim et al. 2007). However, also controversial results have been found. This is due to the fact that condom use is more common in non-marital sexual relations (Kalichman et al. 2007; Johnson e al. 2017). Thus, it is hard to predict what kind of results this factor will reveal in the logistic regression model.

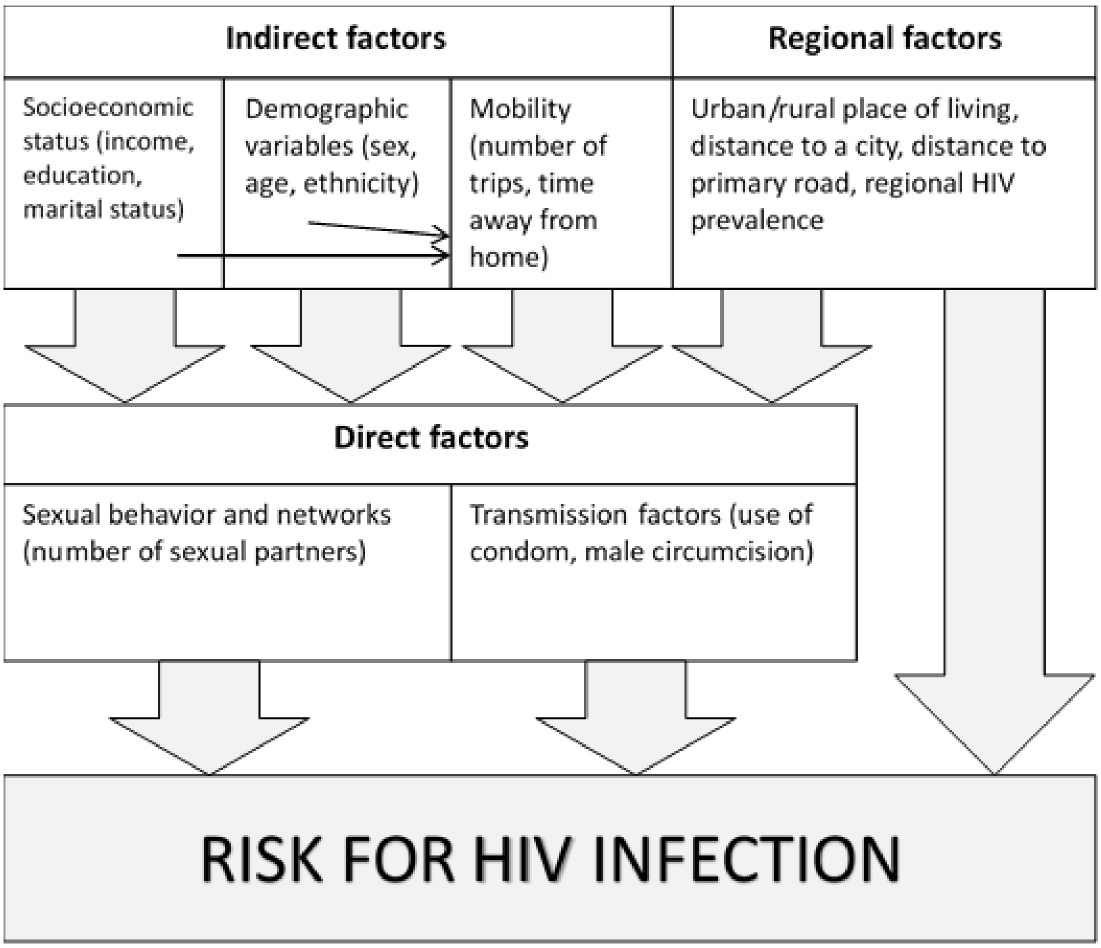

Figure 15. The theoretical framework for the analysis of this study.

4 Data and methods

4.1 Data

4.1.1 Namibia Demographic and Health Survey 2013

The main data used in this study is the Demographic and Health Survey data for Namibia (NDHS) collected in 2013. The DHS program supports survey studies regarding the demographic and health situation in the countries of the Global South. The program is funded by USAID. It provides countries with technical and financial support in conducting of the survey. It also administers sharing the data with academics globally with the intention of contributing to the research concerning demographic and health issues in the Global South. According to the programs website, by 2015 the DHS program has conducted more than 300 surveys in over 90 countries (DHS Program 2017).

In Namibia, the NDHS 2013 is the fourth survey study conducted with support of the program and it has been carried out by the Ministry of Health and Social Services (MoHSS) in collaboration with the Namibia Statistics Agency (NSA), the National Institute of Pathology (NIP) and ICF International. The NDHS 2013 is the first national survey in Namibia to include HIV testing. For this study, the NDHS 2013 data was available as individual person level data set and some key findings have also been obtained from the published final report of the NDHS 2013 (MoHSS and ICF International 2014).

Surveys of the DHS program follow a complex sample design that enables comparisons between administrative areas, between countries and on a temporal scale. The sample design is a stratified two-staged design. The sampling frame used for the NDHS 2013 was the preliminary frame of the 2011 Namibia Population and Housing Census, which contains a list of enumeration areas covering the whole country. The enumeration areas served as counting units for the census. Virtually they were geographical areas: in rural areas villages or group of small villages and in urban areas usually city blocks (MoHSS and ICF International 2014).

In the sample design for NDHS 2013, the country was divided to 26 strata for rural and urban enumeration areas in each of the 13 administrative areas in Namibia. The 554 enumeration areas selected for the survey act as sample clusters in the final data. The sample of clusters have been selected in each stratum separately with a probability that is proportional to the number of ordinary households in the cluster. In the second stage a fixed number of 20 households were selected in each of the 554 clusters regardless of their size. This imposes the use of sampling weights when using the data. The use of sampling weights makes the results of analysis representative at national and regional level (Larmarange 2011; MoHSS and ICF International 2014). The final sample included 11,004 households selected in 554 sample clusters. The interviews were conducted during the time period from May to September 2013. From the selected households 10,165 were found occupied at the time of the interviews and from these 9,849 were interviewed successfully. During the interviews 9,940 women aged 15 to 49, 921 women aged 50 to 64 and 5,271 men aged 15 to 64 were selected as eligible for the individual interviews. From these 9,176 women aged 15 to 49, 842 women aged 50 to 64 and 4,481 men aged 15 to 64 were successfully interviewed (MoHSS and ICF International 2014).

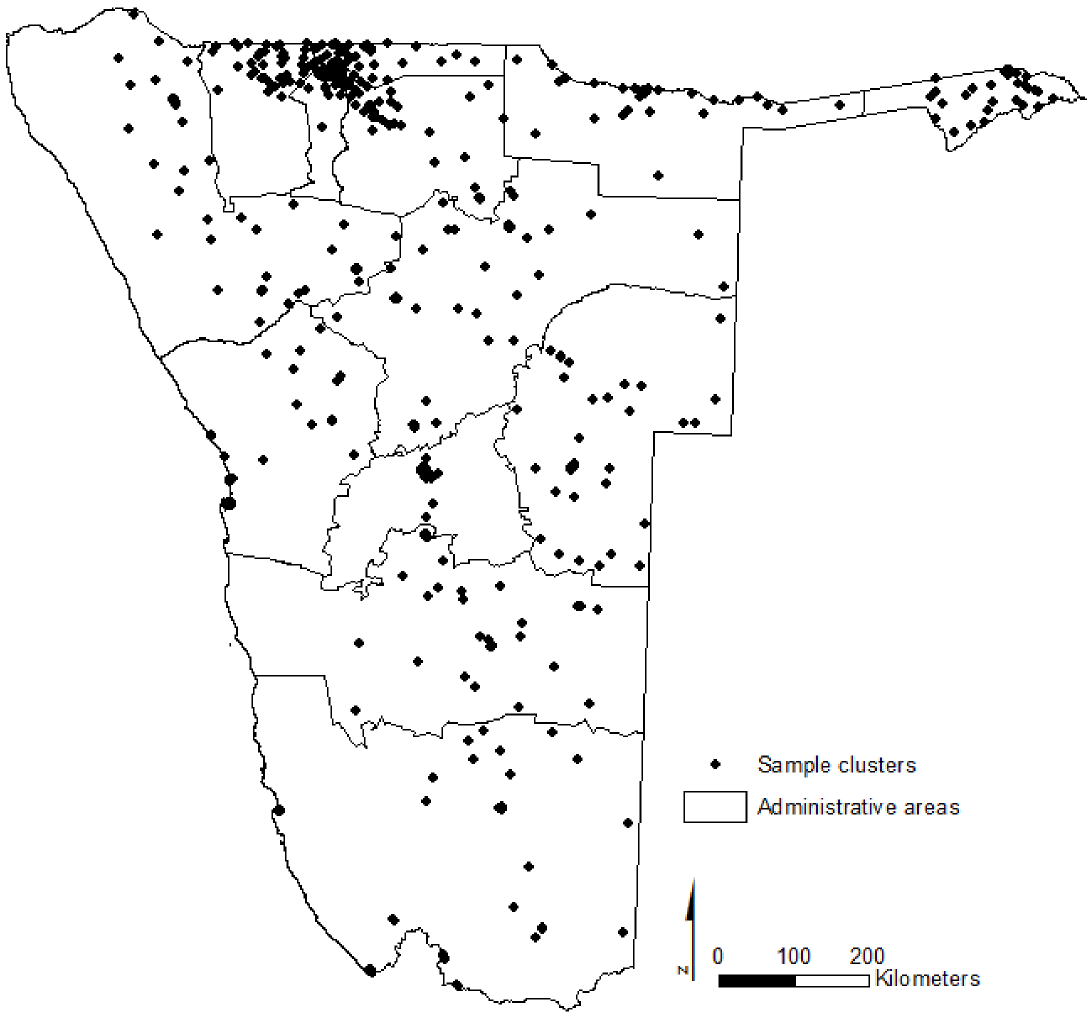

Figure 16. Spatial distribution of the georeferenced sample clusters in the NDHS 2013 data.

The questionnaires include variety of themes from the socioeconomic background information, family planning, attitudes towards HIV and AIDS, health issues and health care to questions about marriage, sexual activity and domestic violence. Furthermore, haemoglobin measurement and HIV testing were done in half of the households that were interviewed. 9,309 individuals were tested for HIV and 8,858 of them were also interviewed individually. From these 8,858 respondents 4,984 were women and 3,874 men. Only those individuals tested for HIV who also participated in the individual interviews were included in the logistic regression analysis.

There are 550 georeferenced sample clusters in the NDHS 2013 data which enables spatial examination of the HIV data. The spatial reference has been collected on the field by the field staff by using GPS devices. The coordinates provided in the data represent the community centroid of the enumeration area which corresponds to the sample cluster. In reality, the respondents of the survey reside in an area surrounding the geographical location of the cluster point. The coordinates for the sample clusters have also been misplaced in a specific manner due to privacy purposes so that the exact location of the respondents is not revealed. The new location for the clusters has been calculated by selecting first randomly a direction between 1 to 360 degrees and after that a random distance between 0 to 2 kilometres for urban clusters and 0 to 5 kilometres for rural clusters. It has also been ensured that after the misplacement the location of each cluster stays inside the correct administrative region and inside the borders of Namibia (MoHSS and ICF International 2013).

The spatial distribution of the georeferenced sample clusters is presented in figure 16. The locations of the sample clusters were available for this study as a complete dataset in Esri shapefile format. The location of the clusters represents the population distribution in Namibia as the sample of clusters were selected separately in each stratum. For valid use of the data, two kind of statistical weights are provided in the data. The first ones are calculated for the use of the whole individual scale data and the other ones for those tested for HIV. In this study the second weights have been used in producing the interpolated sub-regional HIV prevalence estimates and for the logistic regression model. The weights are not designed for the exact sample used in this study because not all respondents tested for HIV participated in interviews other than household interview. However, these weights are more suitable for this data than the weights designed for the whole sample of respondents.

4.1.2 Migration report for the Population and Housing census 2011

In addition to the NDHS 2013 data, some variables for the logistic regression analysis were derived from the Population and Housing census 2011 by Namibia Statistics Agency (NSA 2013). This information has been mainly used in the earlier part of this study where regional socioeconomic situation in Namibia was presented. For this study, final report for the Population and Housing census 2011 (NSA 2013) as well as separate migration report for the census (NSA 2015) were used. The reports provided information on constitutional scale, since the census data was not available for this study on individual person scale.

In the Population and Housing census respondents were asked their birthplace, current usual place of residence, which is where the respondent spends more than six months of the year, duration of residency in this place and usual place of residence in September 2010. With this information, it was possible to review indirectly regional migration rates by counting the difference between the amount of people living in a specific area and the amount of people born in that same area. The difference would then be divided by the people born in that area to get the net implied migration rate for each regional unit. Migration rate was used in this study to examine population mobility on spatial scale. The data was provided in the migration report (NSA 2015) on regional and constituency scale.

Implied migration rate was used in the logistic regression model to indicate regional migration level of the area where the cluster of the respondent is located in. There exists some source of error due to uncertainties in the georeferenced of the clusters which are addressed later on. Clusters were connected with corresponding constituencies according to geographical location using Quantum GIS software.

4.1.3 OpenStreetMap data of primary road network, cities and towns

Primary road network and location of cities and towns in Namibia were obtained from OpenStreetMap project (OpenStreetMap 2016). The data was downloaded from the project webpage in Esri shapefile data format. The data included also attribute data that described the size of the cities and towns as well as road type classifications for the road network. For the analysis of this study, the road network was limited to roads which were classified as primary roads. These are assumed to be important and heavily used main roads that are used for example long-distance travel. For cities and towns, only urban locations classified as cities were chosen for the logistic regression analysis.

4.1.4 Preparation of the data

For the use of the analysis, the individual scale survey data and information about HIV status of the respondents were derived from NDHS 2013 data. Recoded data files were available for download upon request from the website of the DHS program. For the interpolation of sub-regional HIV prevalence, the HIV recode of the NDHS 2013 was used. This data included the results of HIV testing from 9,309 survey respondents. A weight variable designed for the HIV recode of the NDHS 2013 was used when sub-regional HIV prevalence was calculated.

The recode file for men was appended to the women's recode file to create a dataset with all of the 14,499 respondents of the individual interviews. The HIV recode file was then merged with the dataset according to unique key variables appearing in all of the files. Finally, all cases with HIV status missing were excluded from the data file resulting in 8,858 individuals. These were included in the logistic regression model.

For the logistic regression analysis, a merged dataset was compiled from different data sources described in the previous chapters. The scale of observation for the logistic regression model was individual respondent scale. Variables from sources other than NDHS dataset needed to be georeferenced to correspond to NDHS survey clusters according to their spatial location. For example, distance from nearest primary road according to primary road network obtained from OpenStreetMap website was calculated for each cluster point using Quantum GIS software. Distance from cities was calculated in the same manner. The regional migration rates were available on the constituency scale. Migration rates were defined for each survey cluster according to which constituency the survey cluster was located in. From the survey cluster this information was matched to each respondent. The data preparation and analysis were conducted with R-statistics software.

4.1.5 Reliability and validity of the data

It is important to take into account vulnerabilities that exist for the data used in this study. Reliability refers to whether the results from the data are reliable and do not reflect any kind of error. Validity of the data on the other hand refers to whether the methods used to collect the data and conduct the analysis are measuring what they were supposed to. In the following part, sources of error for the DHS datasets are presented, including problems that arise when data from different sources and different geographical scale are combined.

The DHS studies follow a carefully designed sampling method that aims to produce nationally representative samples and results that allow comparison between different countries and in respect to time. The samples are expected to be representative regionally and nationally. Nevertheless, it needs to be acknowledged that in different countries different operators conduct the studies in different circumstances. Even though the sampling methods introduced in the DHS program are highly developed, conclusions should be made with caution when different countries are compared with each other.

More specifically in respect to HIV testing, some studies have argued that the DHS studies including HIV testing most likely contain some non-response bias (Garciá-Calleja et al. 2006; Marston et al. 2008). This derives from the fact that the respondents who denied the HIV testing might be more likely to be HIV positive and hence decline the test. For example, a study that examined non-response bias in nine DHS studies discovered that non-response corrected estimates for HIV prevalence were higher in all the countries (Marston et al. 2008). However, the differences were not statistically significant in all the countries. The study also pointed out that the differences between the HIV prevalence reported by DHS studies and the non-response corrected prevalence were only marginal, usually around one percentage point. The risk for non-response bias in predicted HIV prevalence still needs to be taken into account and for example in 2005 in Kenya a separate HIV and AIDS report was published with the DHS report that included evaluation of the non-response bias (Ministry of Health, Kenya 2005). The evaluation compared the socioeconomic characteristics for the whole survey sample and the non-respondents and found no apparent differences. Nevertheless, it is still important that these matters are acknowledged and source of non-response bias is considered.

In addition to non-response bias, laboratory error needs to be taken into account when using any results from HIV testing. The effect this error might have on the predicted HIV prevalence is most likely marginal. Laboratory error is an issue that is more significant when data from multiple different countries with different conditions and working practices are compared (Garciá-Calleja et al. 2006).

Another issue considering the DHS data and their use in this study are the problems regarding the georeference of the survey cluster points. Firstly, the cluster points only represent the centroid of the community, usually rural village or city block, where the interviews and HIV testing has been conducted. In reality, the people interviewed for the survey reside in an area of indefinable size around this point. Usually this area is not very large but it varies between cluster points (MoHSS and ICF International 2013). This should be acknowledged when using georeferenced data. Another problem with the georeference of the survey cluster points is the deliberate misplacement of the coordinates in order to protect the privacy of the survey respondents (MoHSS and ICF International 2013). The logic behind the misplacement has been explained earlier in this study. The misplacement of the survey cluster points should be kept in mind when spatial analysis is conducted.

Because of the problems in the georeference precise distance from cluster points to roads and urban centres should not be interpreted directly. Such calculated distance does not represent the actual distance from the cluster point due to sources of error in georeferenced. For example, the distance to nearest primary road for two individuals residing around the same survey cluster may in reality be quite different even though the data has same values of distance for these respondents. When this distance is used as an explaining factor for a logistic regression model it should be taken into account that this kind of bias exists in the data. Keeping in mind vulnerabilities of this approach, distances like this have been used in the logistic regression model for lack of more accurate data. Similarly, distances to primary roads and nearest cities have been calculated for survey clusters and individuals also in other studies using the georeferenced DHS data (for example Messina et al. 2010).

Source of error also occurs when data with different spatial scale are combined. This has been done for the implied migration rate data from the Migration Report of the 2011 census (NSA 2015). Since this data was available only in constitutional scale, only average values from these regional units could be used in the analysis. The average migration rate value of the constituency, which the cluster point was located in, was assigned for each cluster point. The constituencies vary in size and the migration rate values assigned to the cluster points give in reality very crude estimations of the migration in-flow or out-flow in the area the respondent resides in. This error could only be limited with more accurate smaller scale data which is difficult to obtain. In addition, the misplacement of the clusters explained earlier causes error when combining the constituency scale data with cluster points. The misplacement is done so that the cluster points still stay inside the correct administrative region. However, the constituency borders have not been considered. With even small error in location, a survey cluster could end up with a migration rate value that is very different from the actual migration rate experienced in the location of the respondent.

In addition to problems with georeference in the DHS data, also the street and city data from OpenStreetMap project include some source of error. OpenStreetMap is open source data that any person can update. It also needs to be acknowledged that some changes may have occurred for example in the road network since the data of the road network was downloaded couple years after the collecting of the survey data.

All these sources of error taken into account, the results of the analysis in this study should be interpreted with caution. However, more accurate spatial data on scale different than NDHS 2013 cluster points was not available to include in the analysis. Including some regional factors to the individual person scale logistic regression model gives the model also a geographical dimension which adds to the previous applications of DHS data in HIV studies.

4.2 Methods

4.2.1 Spatial autocorrelation with Moran's I

The global and local spatial autocorrelation of the HIV prevalence in survey clusters was tested using an open source GIS software GeoDa (GeoDa 2017). Global autocorrelation was tested with Moran's I, which reports how dispersed or clustered the observations in the data are according to the variable of interest. In practical terms, the test measures weather a variable Y at location i is correlated with the same variable at a neighboring location j. In order to do this, the test creates a spatially lagged variable YL. The spatially lagged variable is defined as an average of the values of Y that surround Y at location i. With the help of the spatially lagged variable it is possible to test spatial dependency (Moran 1950).

The neighboring location *j* of location *i* has to be determined with spatial weight matrix. The spatial weight matrix describes the neighborhood structure of the geographical units, in the case of this study the cluster points. The matrix *W* is defined as follows:

$$W_{ixj} = \begin{bmatrix} W_{11} \cdots W_{1\mathrm{j}} \\ \vdots \quad \ddots \quad \vdots \\ W_{i1} \cdots W_{ij} \end{bmatrix}$$

, where *Wij* represents each observation in the data. In this analysis the weight matrix was calculated so that for all survey clusters a fixed stress hold distance was used which determined the neighbours for each observation. This distance was set to 107,934 metres. Histogram of the number of neighbours for the survey clusters is presented in figure 17.

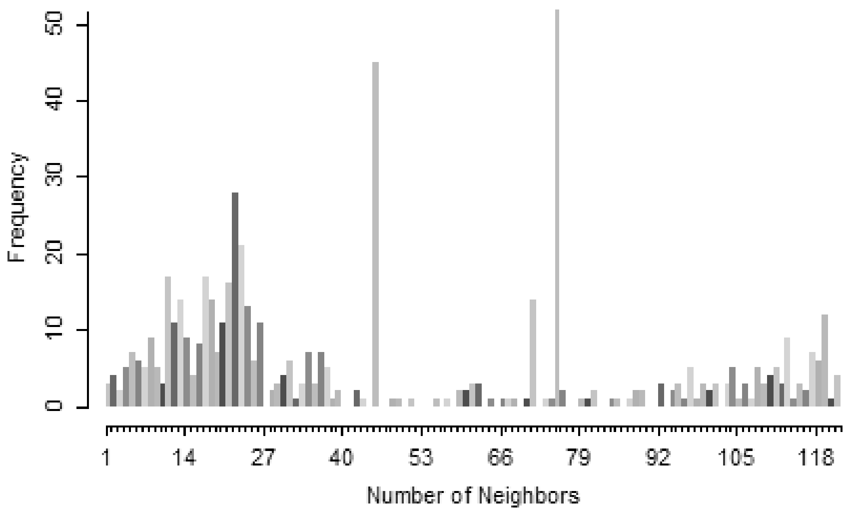

Figure 17. Histogram of the number of neighbours for the survey clusters.

The Moran's I estimates the correlation between the variable of interest *Y* and its spatially lagged counterpart *YL*. Statistical significance of the dependency is estimated by a pseudo p-value. The pseudo p-value is assessed by examining the differences between the original data and a selected number of permutations of the observations. In this analysis, 999 permutations were used to calculate the pseudo p-value.

The local spatial autocorrelation identifies significant clusters of high and low values of HIV prevalence in specific parts of the study area. The local autocorrelation has been assessed in this report by the local indicators of spatial association (LISA) (Anselin 1995). The significance of the specific clusters is estimated with the same procedure as with the global Moran's I. In this report 999 permutations were used to calculate the p-value.

4.2.2 Interpolating HIV prevalence with Kernel density method

A sub-regional smoothed representation of HIV prevalence was generated for Namibia using interpolation. Interpolation methods calculate a continuous surface describing a phenomenon according to known values for scattered points. These methods can be used to predict the spatial distribution of spatially continuous phenomena and they are based on the assumption that points near to each other resemble each other in terms of the studied phenomenon (Tobler 1970). In this study, the interpolation was done with adapted Kernel density method introduced especially for DHS data and HIV prevalence by Larmarange et al. (2011).

Kernel estimator approach generates a continuous surface of the phenomenon in a way similar to the interpolation methods. Intensity surface for the phenomenon studied is created so that different density surfaces are calculated for each observation point assuming that the density decreases with distance. These layers are then combined to generate an overall picture of the phenomenon in the chosen geographical area (Larmarange et al. 2011). The formula for calculating the intensity ŝ of the phenomenon in point (x, y) is as follows:

$$\hat{s}(x,y) = \sum_{i=1}^{n} \frac{1}{h_i^2} K(\frac{d_i}{h_i})$$

, where n is the number of observations, d_i the distance between case of observation i and point (x,y), K a Gaussian Kernel density function and h_i the bandwidth used for i (Larmarange et al. 2011). The bandwidth chosen for the formula determines the level of smoothing of the data and it can be fixed or adaptive (Larmarange et al. 2011).

In the case of DHS data, Larmarange et al. (2011) have used an adaptive bandwidth where the size of the bandwidth depends on the number of cluster points and more precisely on the number of people tested for HIV surrounding the point that the density surface is generated for. A minimum number of observations required in order to calculate the density surface (N) is set, which determines the size of the bandwidth used. In other words, where clusters have fewer observations and clusters are further apart bigger bandwidths are used whereas smaller bandwidths are used when clusters are close to each other and include more cases of observation.

The HIV prevalence is defined for each survey cluster by dividing the number of HIV positive respondents with the total number of respondents tested in the cluster and then multiplied by hundred. Hence, it describes the number of HIV positive respondents by 100 respondents in total. The minimum *N* which determines the bandwidth size was set to 300 tested individuals for the total sample and 200 when men and women were examined separately. For the computation of the interpolated raster surface also weights provided in the NDHS 2013 HIV module were used.

In this study, the adapted Kernel density method has been applied to NDHS 2013 HIV data using R statistical software and prevR package programmed by Larmarange (2013). The package provided a user-friendly way of utilising the adapted Kernel density method for HIV data from DHS studies. The package calculated the prevalence surface from the input data with the preferred options. The file format was then changed so that the generated surface was suitable to export to GIS software. The final raster surfaces were then visualised in Quantum GIS software.

One key objective of this study was to produce interpolated sub-regional representations of HIV prevalence in Namibia. The variation in the intensity of the HIV epidemic has been examined through visual interpretation. HIV prevalence maps have also been generated for men and women separately to highlight the differences between them. Larger scale maps for North-Central Namibia and Windhoek have been produced to examine these areas more closely.

For the logistic regression model, the average HIV prevalence inside a 25-kilometre buffer surrounding a cluster point was used as an independent variable. The average HIV prevalence inside this buffer was calculated using the HIV prevalence interpolated with Kernel density method (Larmarange 2013). This was done using the Quantum GIS software. A similar approach has also been adapted by another HIV study using logistic regression model and DHS data (Messina et al. 2010).

4.2.3 Logistic regression model

In this study, the influence of variety of factors to the individual person scale HIV risk was examined using logistic regression model. Aim was to examine dependencies between HIV status and selected demographic, socioeconomic and regional factors as well as factors related to the sexual behaviour of the respondent. Three different models were compiled, one for all the respondents, one for women and one for men.

Logistic regression is a modification of regression analysis and it is generally used when categorical or binary variables are examined as dependent variable. Logistic regression analysis predicts the risk of realisation of the phenomenon studied, in other words realisation of the dependent variable, depending on a selection of independent variables. Logistic regression is defined by the following formula:

$$\text{logit} = \ln[\frac{P(Y=1)}{1-P(Y=1)}] = a + bx$$

, where P(Y=1) is the probability that the phenomenon studied will realise and 1-P(Y=1) is the probability that the phenomenon will not realise. In other words, compared to general regression formula, in logistic regression the dependent variable is replaced with natural logarithm of the probability that the phenomenon studied will realise divided by the probability that the phenomenon will not realise. This ratio is also called the odds ratio. From the model parameter (*logit*) odds ratio can be calculated with the inverse exponential function of natural logarithm:

$$\text{OR} = e^{logit}.$$

Odds ratio discloses how many times higher the risk for the phenomenon studied is to realise when the value of the independent factor is increased by one or in the case of a categorical independent factor, when a case of observation belongs to a specific category in relation to a chosen reference category. This enables examination of the dependencies between binary dependent variable and nominal categorical independent variable and reveals the statistical significance of the relationship. Logistic regression model is a common method in population and health studies. Especially in HIV studies the method is useful because binary variable of HIV status for individual is predicted using categorical or ordinal independent variables.

For this study, variables from the NDHS data were used as independent factors. These include sociodemographic factors such as sex, age, educational attainment, wealth quintile, marital status and language group. Also, the urban versus rural status of the survey cluster was used in the model as an independent factor. Individual's population mobility was indicated with number of trips one had done during the last 12 months. Additionally, regional implied migration rate was used to indicate the influence of population mobility level in the area regardless of respondent's own migration status. In the analysis, also sexual behaviour of respondents was used to estimate HIV risk. Variables that were chosen to describe sexual behaviour were number of sexual partners during respondent's lifetime and for men whether respondent used condom during the last sexual intercourse.

The final selection of independent variables was chosen after testing multiple candidate models with different combinations of variables. To estimate the model performance, several estimates for the goodness of fit were used. One of these was Akaike's Information Criterion method (AIC). This method gives an estimate of the relative quality of the model. The AIC can be used especially to make comparisons between different models and estimate the statistical probability of the better performing model to reduce information loss in the model (Akaike 1974).

In addition to AIC, the predictive ability of the candidate models was assessed to choose between available assemblies of independent variables. The predictive ability was tested using part of the data in the fitting of the model and another part to test the parameters the first part of the data produced. The data was divided randomly to these two parts so that the first part covered 75 percent of the observations and the last part 25 percent. Assessment of the predictive ability was done by setting a cut point for estimated probability to 0.5 for the dichotomous dependent variable (Hosmer et al. 2013).

Model performance, and more specifically classification accuracy, can also be assessed with the area under the Receiver Operating Characteristics (ROC) curve. The ROC curve can be plotted and the area under the curve calculated to assess model performance. The AUC values can vary from 0.5 to 1. The higher the value the better the classification accuracy (Hosmer et al. 2013; Turner 2013). This measure is currently the most common way of assessing classification accuracy for logistic regression model. The advantage of the measure is that it does not base on single cut point for the estimated probability but instead plots the prediction accuracy for all possible cut points (Hosmer et al. 2013).

The use of sample weights in a logistic regression model shares opinions. When the sampling design is complex and sample weights are available, their use is recommended especially when estimating simple parameters such as statistical average or relative frequencies. Nevertheless, in more complex statistical modeling, the use of sample weights has not been unambiguous, although many studies have shown that some source of error can occur when sample weights are not used (Cambless and Boyle 1985). In this study, the sampling weights provided in the NDHS 2013 data for the HIV module were used in the logistic regression model.

5 Results

5.1 Visual interpretation of spatial variation in HIV prevalence

One of the key objectives of this study was to produce representations of HIV prevalence distribution on different geographical scales of observation. Earlier in this study, the HIV prevalence in Namibia was presented on national scale as well as on administrative region scale. In the next part, the scale of observation has been focused to produce representations that describe sub-regional differences in HIV prevalence. In the first part, HIV prevalence has been presented on survey cluster scale and in the following part as interpolated sub-regional estimates. The latter form a smoothed surface that estimates HIV prevalence also in locations where direct measurements regarding the phenomenon have not been made.

5.1.1 HIV prevalence on NDHS 2013 survey cluster scale

In the following examinations HIV prevalence is examined in NDHS survey clusters. The HIV prevalence represents the number of HIV positive respondents per 100 people tested in the survey cluster. The observed values for HIV prevalence in cluster points are highest in the North-Central Namibia and the Caprivi strip (figure 18). The same areas have high HIV prevalence also when examined on regional scale (figure 14).

Concentrations of higher HIV prevalence seem to occur where there are bigger cities in Namibia. When comparing the largest concentrations of high HIV prevalence to the map of cities and towns in figure 4 it can be noted that a big part of these concentrations outside the North-Central Namibia and Caprivi are located where there are large cities or towns. For example, in the proximity of Windhoek, Swakopmund, Walvis Bay, Mariental and Lüderitz there seems to exist survey clusters with high HIV prevalence. Outside of densely inhabited North-Central Namibia and Caprivi the clusters with high HIV prevalence can be found especially in urban areas. However, also highest population densities can be found in these areas (figure 4).

In figure 18, North-Central Namibia and Windhoek have been examined more closely. The spatial distribution of the survey cluster points crudely represents the distribution of population in these areas. Crudely it would seem that in North-Central Namibia the single points with highest values of HIV prevalence exist where there are a lot of survey clusters and hence population. The cluster points, which are located outside of this most inhabited area, seem to have lower HIV prevalence. The pattern is the same in Windhoek where most cluster points are located in the northern constituencies. In these areas also exist the clusters with the highest HIV prevalence. In fact, in the southern parts of Windhoek there seems to be significantly lower HIV prevalence in all the clusters when compared to northern parts of the administrative region.

According to the histogram and cumulative frequency of HIV prevalence values in surveys clusters (figure 19), notable part of the survey clusters has a HIV prevalence of zero. Clusters with no HIV positive respondents cover 25 percent of all the survey clusters. Nevertheless, a large part of the survey clusters still has quite high HIV prevalence. The highest quartile of clusters has HIV prevalence higher than 20 percent. Median for the HIV prevalence in all the survey clusters is 11.3 percent and the mean value for HIV prevalence is 13.3 percent. There is a slight difference in the mean value for HIV prevalence reported here to the mean HIV prevalence presented in the final report of the NDHS 2013, which is 14 percent. This small difference derives for example from the fact that in this study only respondents both tested for HIV and interviewed in the NDHS are included. This sums up to 8,858 respondents when HIV prevalence estimates in the NDHS final report have been calculated using the whole population tested for HIV which accounts 9,309 respondents. However, both estimates have been calculated using sample weights provided in the NDHS data.

HIV prevalence in survey clusters is also examined here separately for both sexes. For these examinations, the original individual scale data was divided by sex and the HIV prevalence in each cluster was calculated only according to this population. According to these statistical examinations, men seem to have significantly lower HIV prevalence in survey clusters than women (figures 21 and 23). For men, the mean value of HIV prevalence in survey clusters is 10.7 percent when the corresponding value for women is 15.6 percent. It needs to be noticed, that when the data is split according to sex the total count of respondents for single clusters decrease and hence more zero values of HIV prevalence occur. Especially for men, the number of clusters with zero HIV prevalence increases following in half of the clusters having zero HIV prevalence. For women, this share is around 38 percent. Because of the large number of clusters with zero HIV prevalence, the cumulative frequency curve for men does not seem notably steeper than the one for women (figures 21 and 23). However, the maximum value for HIV prevalence for men is around 67 percent which is less than the maximum of 73 percent for the whole sample with both sexes. For women, the maximum HIV prevalence in a cluster is 100 percent. This observation is caused when a small sample of respondents is used to calculate the HIV prevalence for a single cluster.

Based on maps in figures 20 to 22, the overall picture shows that men have less clusters with high HIV prevalence values and more clusters that have low HIV prevalence. For women survey clusters tend to have higher prevalence. It is hard to say according to cluster scale examinations whether the intensities in the geographical distribution of HIV prevalence change when only men or women are taken into account. Nevertheless, some slight differences can be noticed. For example, for men HIV prevalence seems to be higher in clusters located near some of the big cities, for example Rundu and Omuthiya. In areas that experience the highest overall HIV prevalence, including North-Central Namibia and Caprivi, women have many survey clusters with significantly high HIV prevalence. Also in areas around Windhoek, women seem to have more clusters with high HIV prevalence than men.

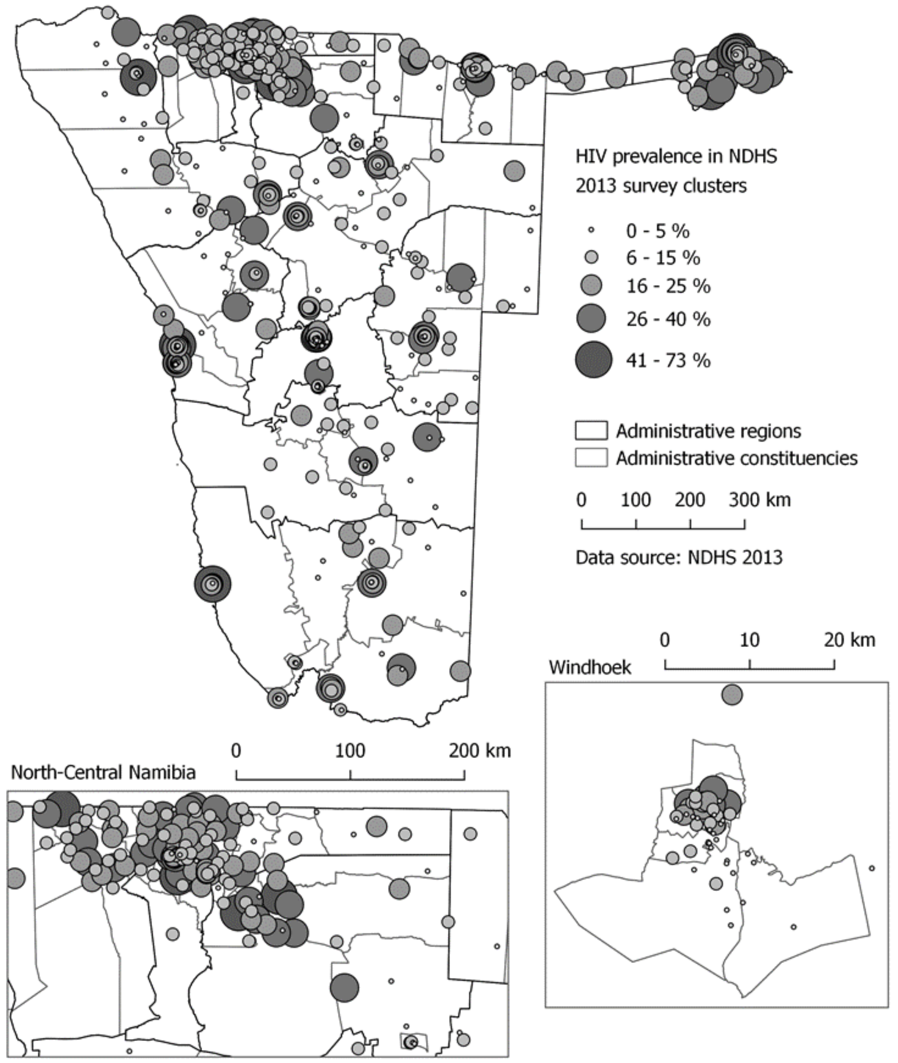

Figure 18. HIV prevalence in the survey clusters of the NDHS 2013.

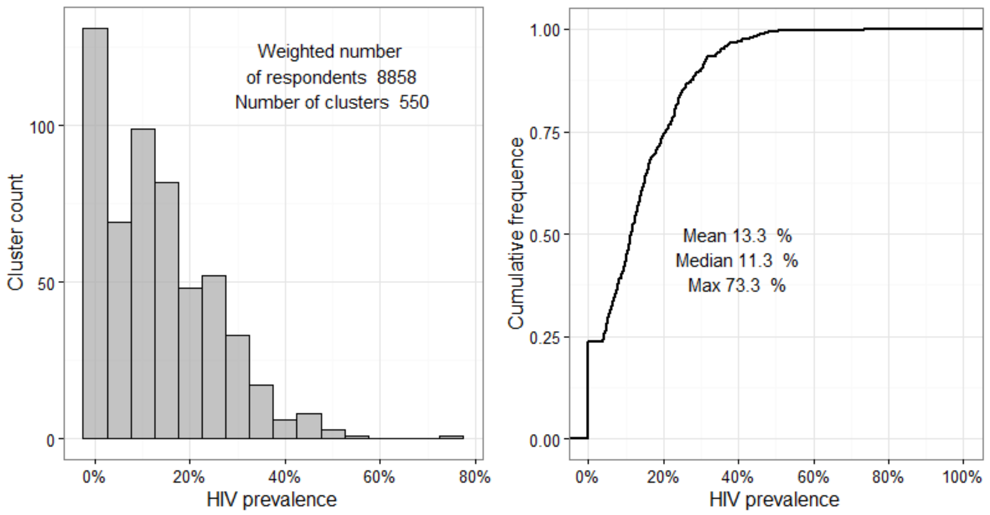

Figure 19. Histogram of HIV prevalence in NDHS 2013 survey clusters.

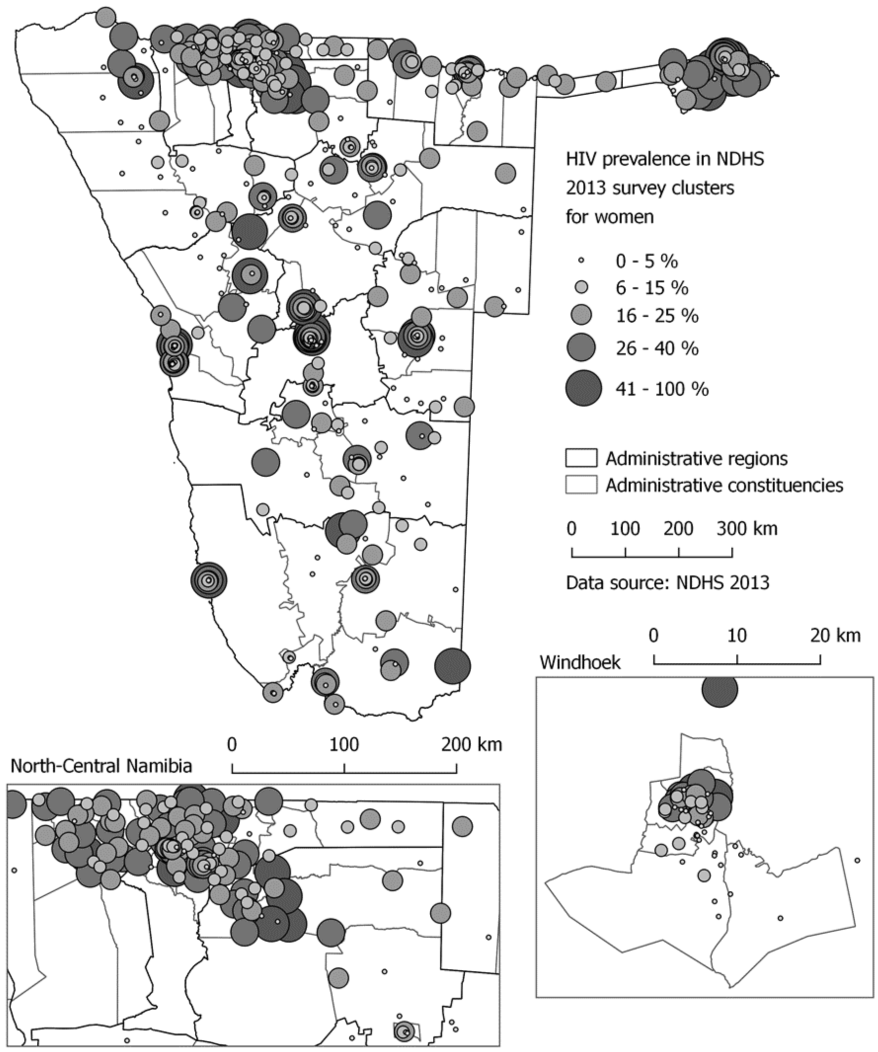

Figure 20. HIV prevalence in the survey clusters of the NDHS 2013 for women only.

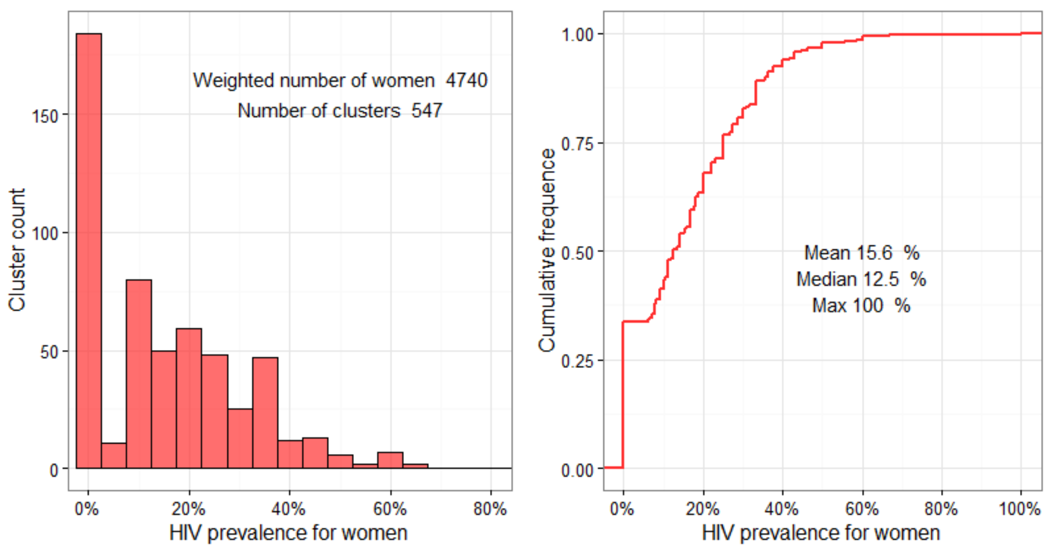

Figure 21. Histogram of HIV prevalence for women in NDHS 2013 survey clusters.

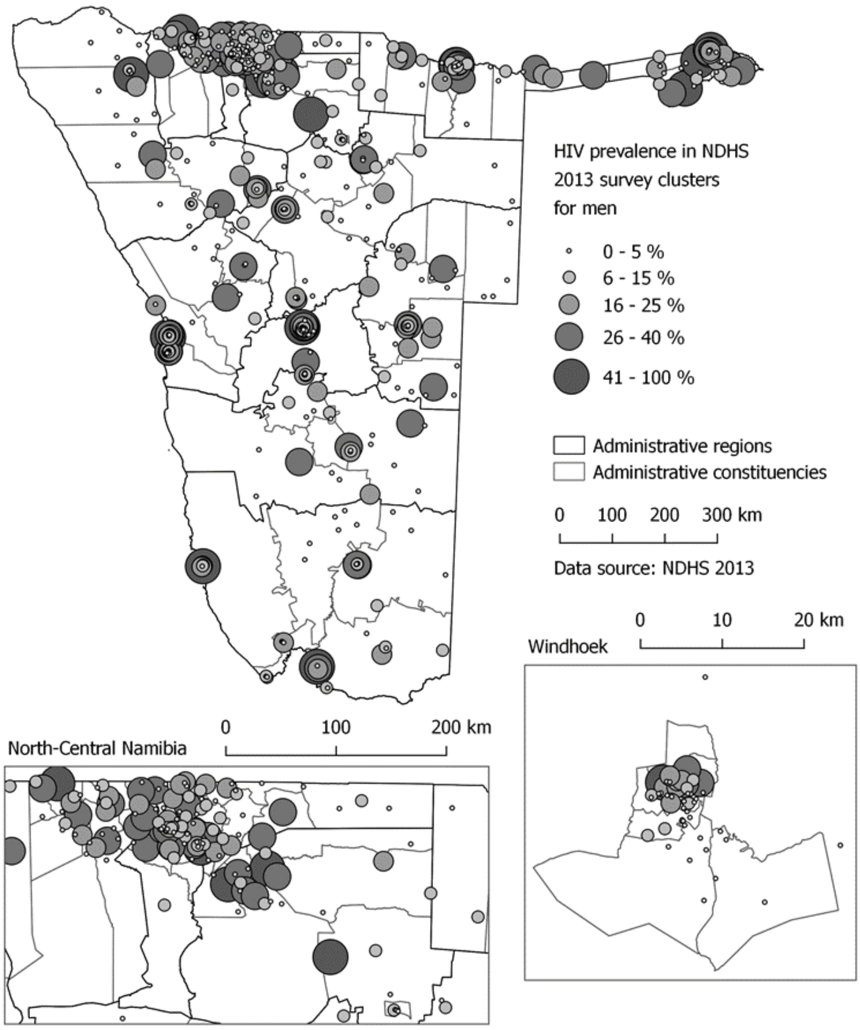

Figure 22. HIV prevalence in the survey clusters of the NDHS 2013 for men only.

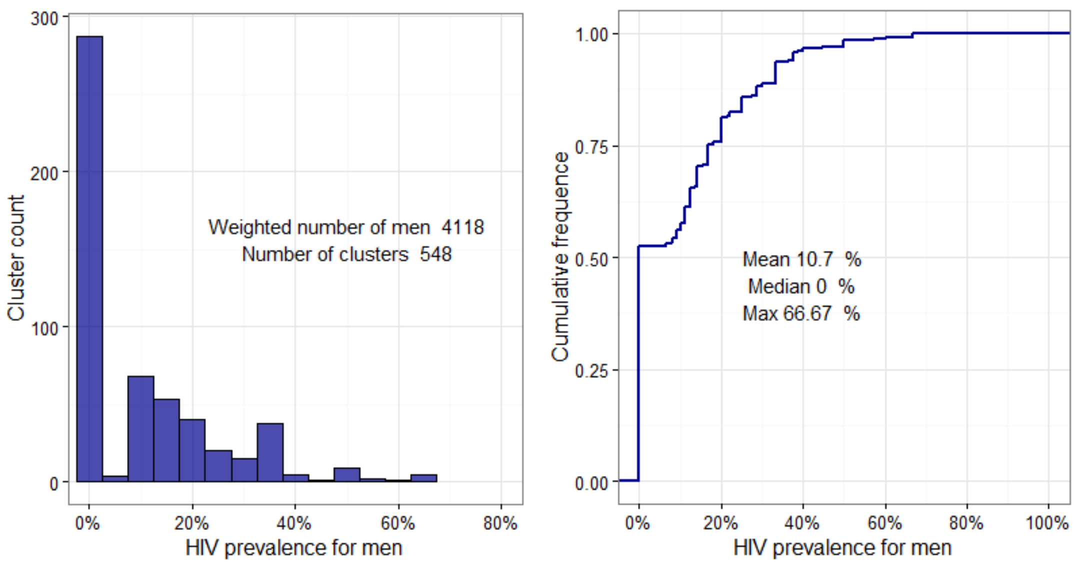

Figure 23. Histogram of HIV prevalence for men in NDHS 2013 survey clusters.

5.1.2 Spatial autocorrelation of HIV prevalence in survey clusters

The spatial nature of HIV prevalence in Namibia was approached in this study by testing the global and local spatial autocorrelation for HIV prevalence in NDHS 2013 survey clusters. According to global Moran's I, a statistically significant correlation was found between HIV prevalence and its spatially lagged counterpart (figure 24). In other words, the HIV prevalence values in neighboring survey clusters correlate with the HIV prevalence in a specific cluster point. This supports the decision of including interpolated HIV prevalence in the final logistic regression model since high and low values of HIV prevalence seem to be concentrating to certain locations.

HIV prevalence experiences autocorrelation especially in North-Central Namibia, Caprivi and Windhoek (figure 25). In North-Central Namibia, there exists statistically significant clustering of high HIV prevalence values. The case is the same for Caprivi. Nevertheless, in North-Central Namibia there seems to exist also survey clusters of low HIV prevalence which have neighbors with high HIV prevalence. In Caprivi, these clusters are not as common.

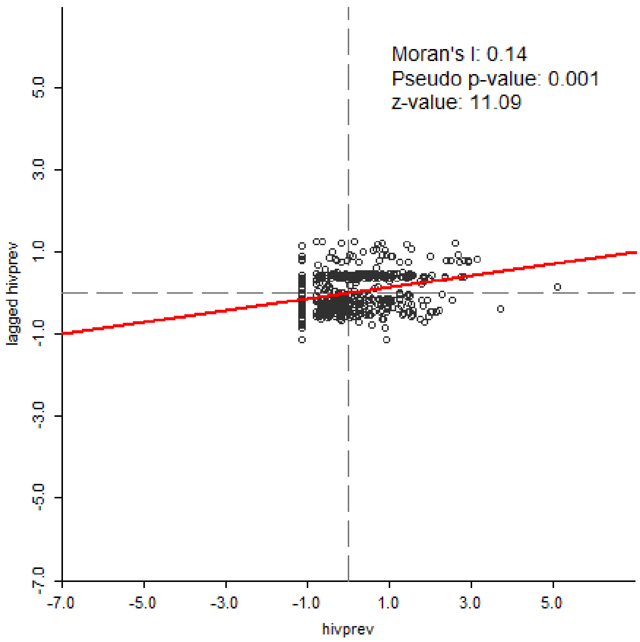

Figure 24. Moran's I scatterplot for HIV prevalence in survey cluster points and the spatially lagged counterpart. The pseudo p-value according to 999 permutations of the data is shown in the graph.

Elsewhere in Namibia clusters of low HIV prevalence seem to be dominating the overall picture. Especially in Windhoek, there seems to be clustering of low HIV prevalence survey clusters. In northern parts of Windhoek survey clusters of high HIV prevalence are neighbored by clusters of low prevalence. It is also evident according to the cluster scale examination of HIV prevalence, that the northern parts of Windhoek experience higher level of HIV prevalence than the southern parts (figure 18).

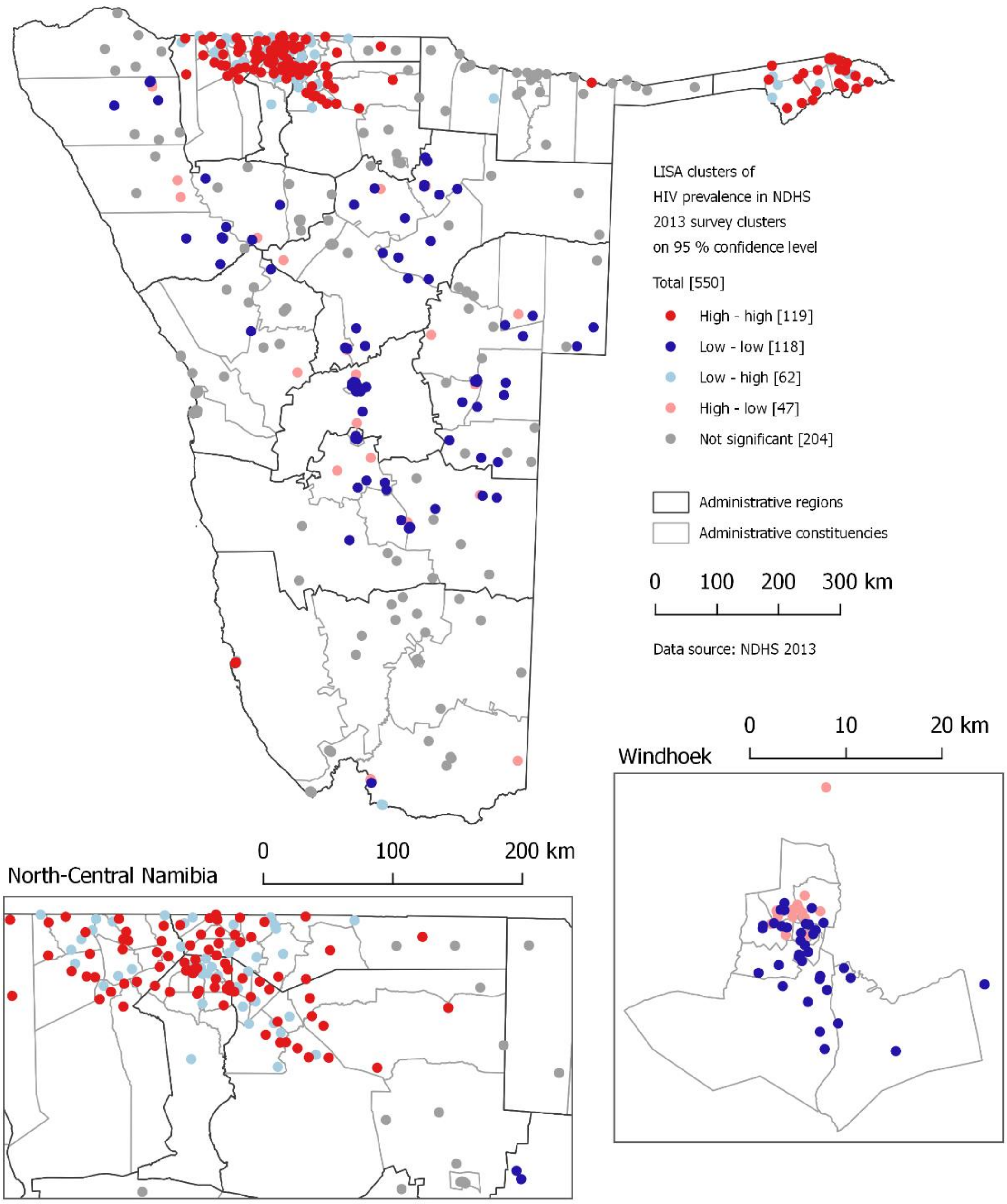

Figure 25. Statistically significant autocorrelation clusters of HIV prevalence in survey cluster points with the pseudo p-value smaller than 0.05. The pseudo p-value was calculated using 999 permutations of the data.

Outside of North-Central Namibia and Caprivi clusters of low HIV prevalence seem to be most dominant but the town of Lüderitz on South-West coast seems to form an exception. In Lüderitz, there seems to exist clustering of high HIV prevalence survey clusters. This is an interesting finding since for example in Swakopmund and Walvis Bay further north on the coast do not experience statistically significant spatial autocorrelation in regard to HIV prevalence even though in these areas HIV prevalence seems to be as high.

5.1.3 Interpolated sub-regional HIV prevalence

Based on the HIV testing conducted in the NDHS 2013 survey clusters, HIV prevalence was interpolated in this study to form a continuous raster surface presenting the estimated HIV prevalence in different areas of Namibia (figures 26 to 28). In figure 26 the raster surface has been calculated taking into account both women and men. Values for HIV prevalence in this representation range from 0.34 to 30.2 percent. The highest HIV prevalence for both sexes can be found in the Caprivi strip, North-Central Namibia and also in areas around large urban centres. These include Rundu in Kavango region, Swakopmund and Walvis Bay on the west coast in Erongo region, Umuthiya and Outapo in North-Central Namibia and parts of Windhoek and its surroundings. It seems that the estimated HIV prevalence is highest where there is also high population density and urban centre present.

Nevertheless, this is not the case for all areas or urban centres. For example, according to Population and Housing Census, city of Lüderitz by the coast in Karas the region with population of little over 12,500 does not have a high population density (figure 5). The city also experiences decreasing in population. However, the area around this city would seem to experience relatively high HIV prevalence and also spatial autocorrelation of survey clusters with high HIV prevalence (figure 25).

In Omaheke region where there is relatively dense population compared to many other areas in Namibia, not notably high values for HIV prevalence are found. In this region is located the city of Gobabis, which has population of 19,000 and growth rate of 37.9 percent. In Hardap region is located city of Mariental with population of 12,500 but also this area has low estimated HIV prevalence. Some urban areas would seem to embody high HIV prevalence while others do not. To some extent, this could be explained with differences in population density but according to these crude observations also urban areas with growing population like for example Gobabis in Omaheke region in some cases have low estimated HIV prevalence.

In areas with high population density, interpolated HIV prevalence estimates have been calculated according to more cluster points in respect to areas with lower population density and fewer cluster points. Due to this, estimated HIV prevalence is not only usually higher in areas of high population density but they also follow patterns that are more elaborate. In North-Central Namibia, population is focused in the northern part of this area (figure 5). Nevertheless, HIV prevalence is highest in the north-western part of this area around the city of Opuwo as well as in the Oshikoto region around the city of Omuthiya and Oshana region in the surroundings of the city of Oshakati. Especially in areas to the south from Omuthiya, population density is not particularly high. According to these visual interpretations, in North-Central Namibia presence of urban centres seems to affect estimated HIV prevalence more than population density by itself. Around some urban centres such as Omuthiya HIV prevalence seems to be higher than others, for example Oshakati, which is located in the northern part of the Oshana region.

In Windhoek, the interpolation produces interesting variation in the estimated HIV prevalence values. According to the representations of HIV prevalence on survey cluster scale (figure 18), the northern constituencies in Windhoek contain clusters with high HIV prevalence. Almost all of the clusters in the southern constituencies have a HIV prevalence of less than 5 percent. This in evidently reflected in the interpolated values of HIV prevalence in this area. In figure 26, the area around Windhoek is clearly delineated to the northern parts with high HIV prevalence and southern parts with low HIV prevalence. The southern parts do not have as high population density as the northern parts of Namibia. Still this area is densely populated and it contains several survey clusters. In other words, there seems to exist steep inequalities in regard to HIV risk inside the area of Windhoek.

In figures 27 and 28, the continuous raster surface for HIV prevalence has been calculated for women and men separately. According to these representations, women have clearly higher HIV prevalence in most parts of Namibia. However, differences in density of interpolated HIV prevalence between women and men are harder to detect by visual interpretation. In other words, it is not hard to say if women have relatively lower HIV prevalence somewhere where men on the other hand have higher HIV prevalence.

Nevertheless, slight differences can be found between these representations. For example, for women high values of HIV prevalence seem to be more spatially concentrated and for men areas with relatively high HIV prevalence are more scattered. Areas between North-Central Namibia and Windhoek have relatively high HIV prevalence when only men are taken into account. For women these areas do not have relatively as high prevalence. The same pattern exists for men also in the north-eastern areas of Kunene region, southern parts of the Omaheke region and the area on the border of Kavango and Caprivi regions. These areas have high HIV prevalence values in relative terms for men when compared to women.

For women areas with relatively high HIV prevalence are for example southern parts of Erongo and Otjozondjupa regions north from Windhoek. For men these areas do not have as high HIV prevalence. In southern parts of Karas region, women also have high HIV prevalence values where men do not. Also south from the city of Rundu inside the area of Kavango and Otjozondjupa regions, women have quite high estimates for HIV prevalence where estimates for men on the other hand are relatively low.

The North-Central Namibia seems to display interesting pattern when women and men are observed separately. Both sexes experience high HIV prevalence values around the city of Omuthiya. However, for women HIV prevalence is very high throughout the area. For men on the other hand high values of HIV prevalence are concentrated almost exclusively around Omuthiya and the very western part of the Omusati region. All the areas around other cities in North-Central Namibia, for example Outapi, Oshakati and Eenhana, display relatively low HIV prevalence estimates for men.

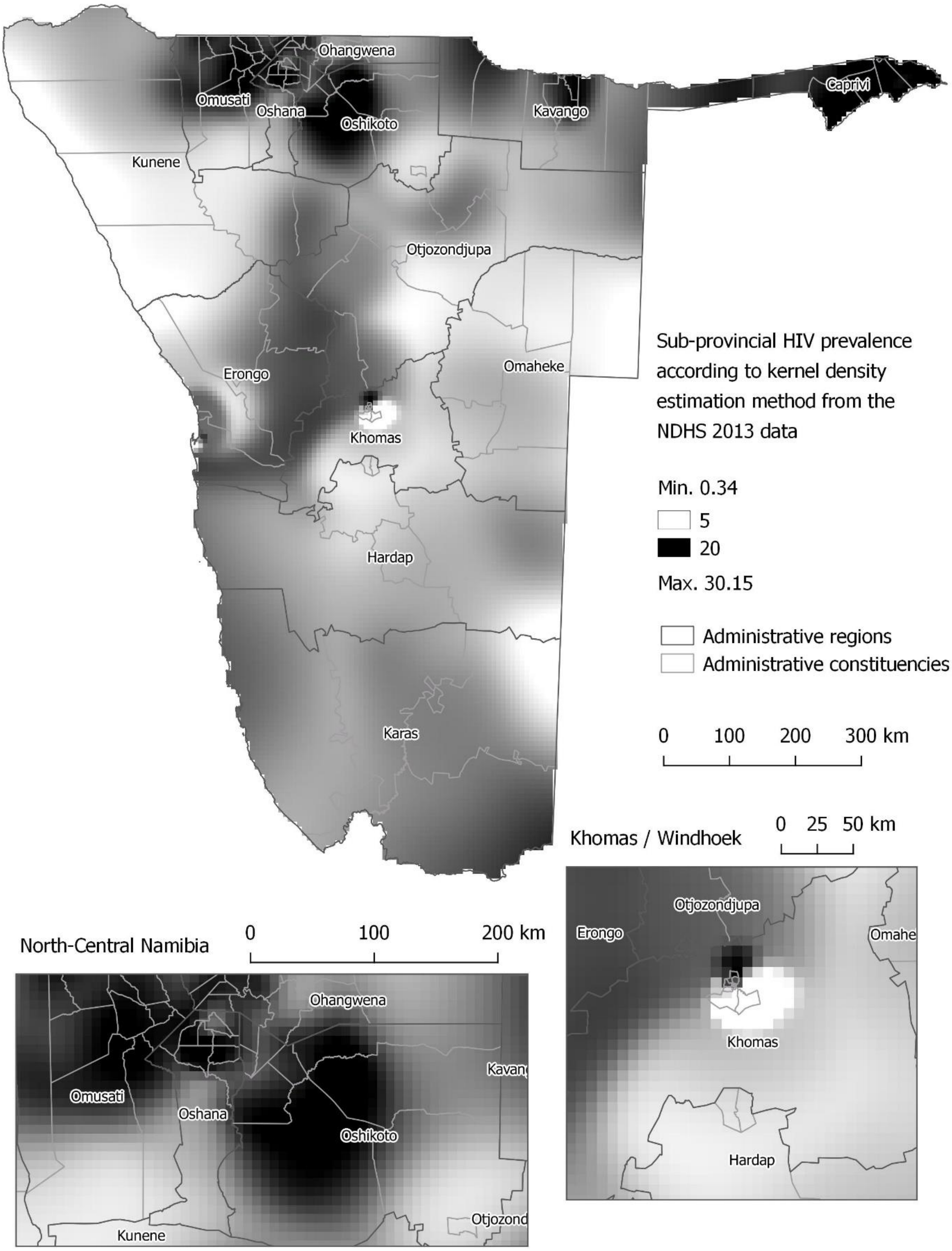

Figure 26. Sub-regional HIV prevalence according to kernel density estimation method from the NDHS 2013 data.

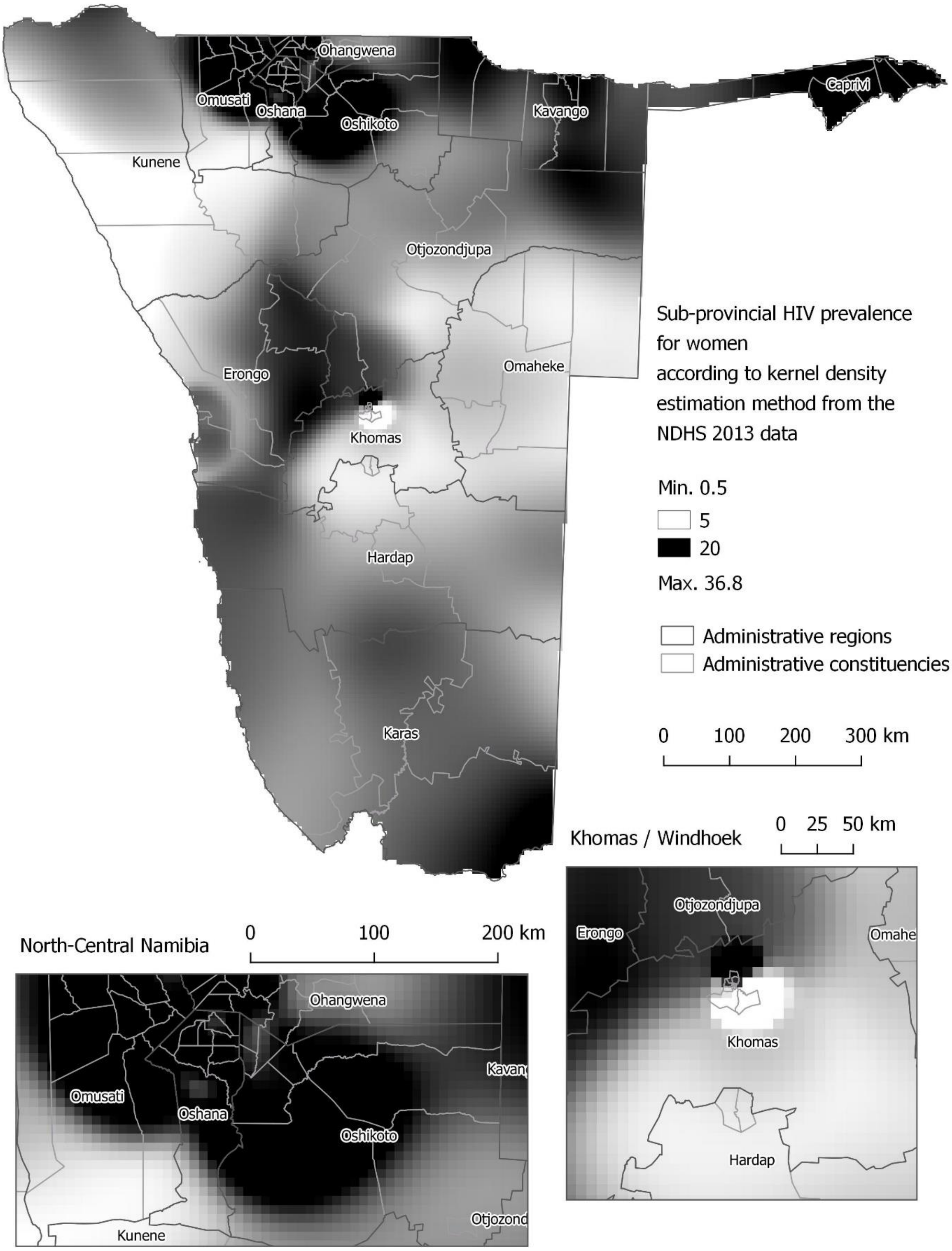

Figure 27. Sub-regional HIV prevalence for women according to kernel density estimation method from the NDHS 2013 data.

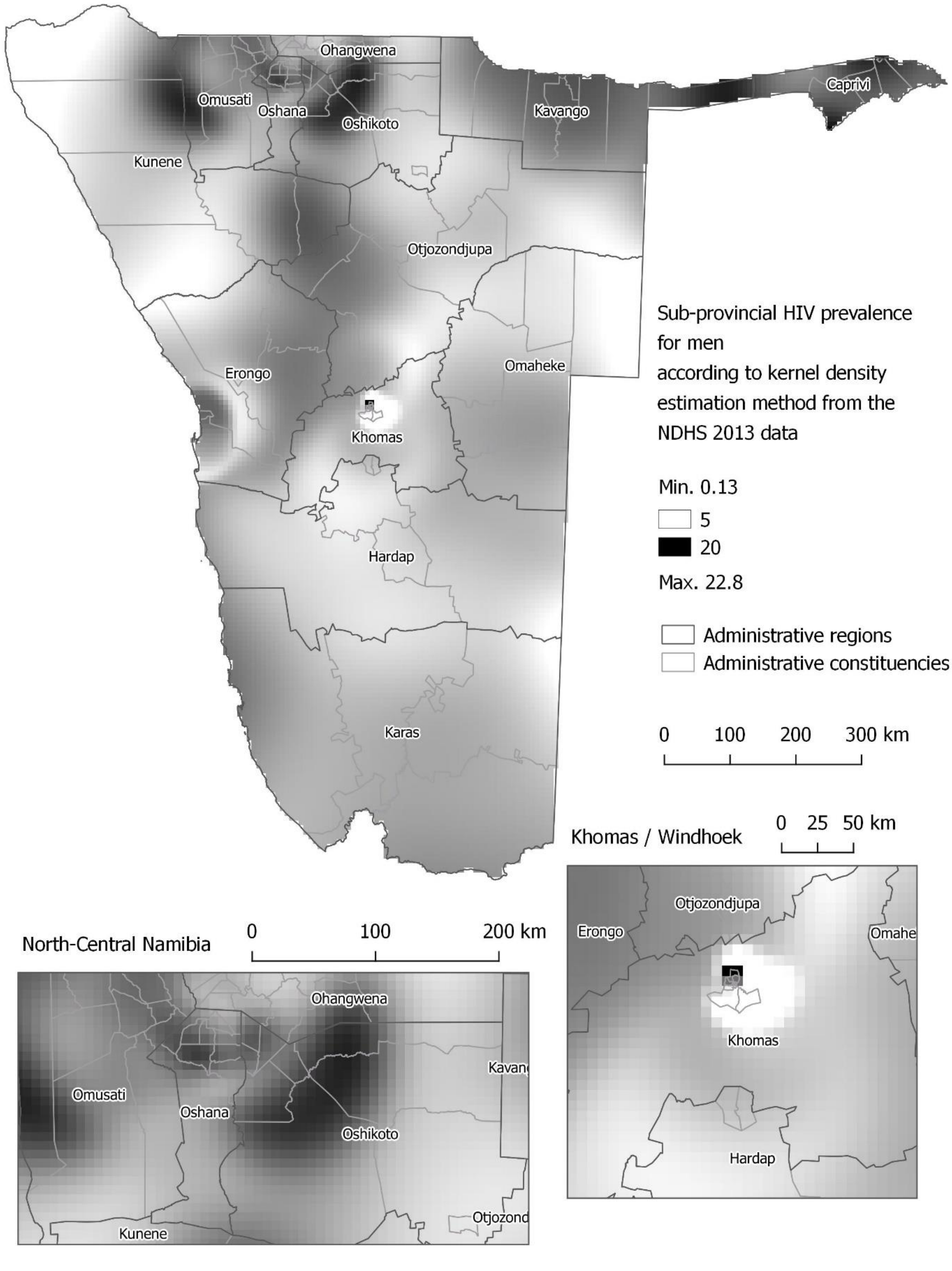

Figure 28. Sub-regional HIV prevalence for men according to kernel density estimation method from the NDHS 2013 data.

5.2 Logistic regression model

5.2.1 Descriptive statistics

In table 3 are presented frequencies of the categorical variables included in the logistic regression model and percentage of HIV positive respondents for each reference group. Sample weights provided in the HIV module of the NDHS 2013 data have been used in this table. The factors chosen for the logistic regression models have been used before in previous studies to explain HIV risk. Some of them were already examined in the theoretical background of this study according to the NDHS 2013 final report (MoHSS and ICF International 2014). These include for example average HIV prevalence for women and men, different age groups (figure 13), wealth quintiles and groups for different level of educational attainment. In the next parts after the descriptive statistics, I present the results of the logistic regression analysis and evaluate model performance.

In the weighted data used for the logistic regression model 1,269 of the total 8,858 respondents were HIV positive. This accounts for around 14 percent of the respondents. Among women the weighted percentage is 17 percent and for men 11 percent (table 3). In the weighted data 3,941 respondents were women and 3,647 were men. From different age groups the percentage HIV positive is highest among those aged 30 to 49. When men and women are examined separately, for men the age group with highest percentage is from 40 to 49 and for women 30 to 39. HIV prevalence is lowest among 15 to 19 year olds of which only two percent are HIV positive.

From different socioeconomic groups the HIV prevalence is highest among those who had no finished educational attainment or primary education or among those who are in the lowest or second wealth quintile. In regard to educational attainment, for women HIV prevalence is highest among those with primary education and lowest among those with higher than secondary education. For men HIV prevalence is highest among those with no educational attainment. In regard to wealth quintiles, for women the highest HIV prevalence can be found in the second wealth quintile where it is 24 percent as well as for men for which the respective percentage is 16. Among different language groups highest percentage of HIV positive respondents can be found among those who reported speaking Lozi or Oshiwambo.

The variables indicating differences in sexual behaviour in the analysis are lifetime number of sexual partners and for men also condom use. The frequencies regarding HIV status do not seem to produce significant differences when number of sexual partners is examined. For men and women, among those who reported having had sex with more than 20 partners during their lifetime the HIV prevalence was highest. Nevertheless, for women the prevalence was also high among those who had 5 to 20 partners. Differences between the reference groups are not wide. For all respondents, as well as for women and men separately, the HIV prevalence is significantly low among those who reported not having had sex during their lifetime.

Among respondents who were registered in urban survey clusters, HIV prevalence is slightly higher than for those registered in rural clusters. For women this difference is a little wider. Among those living closer to a city according to the data adapted from OpenStreetMap project, the percentage of HIV positive respondents seems to be somewhat higher than among those living further away from a city. The same trend can be found when distance to primary road is examined. The differences in the frequencies between the reference groups are not wide and most likely not statistically significant.

When regional migration rate of the constituency of the survey cluster and number of trips respondent reported having done during the last 12 months are compared, highest HIV prevalence can be found among those who had not made any trips during the last 12 months. For women the difference in HIV prevalence between those who had not made any trips and those who had made several were quite wide. In addition, woman respondents reporting to have done only one to four trips were more likely HIV positive than those reporting to have done more trips than this. For men differences between these reference groups were marginal.

In regard to regional migration rates, those who were registered in a survey cluster that is located in a constituency experiencing lifetime migration outflow are more often HIV positive than respondents in other reference groups. For women the reference groups with highest percentage of HIV positive respondents are the ones experiencing outflow. For men respectively, the reference groups with migration inflow had highest HIV prevalence. Nevertheless, differences between reference groups were quite marginal.

Table 3. Frequencies of the categorical variables from NDHS 2013 included in the logistic regression model and percentage of HIV positive respondents in each reference group. Sample weights have been used.

		Women and men					Women					Men				
		Positive		Negative		Total	Positive		Negative		Total	Positive		Negative		Total
Sex	Male	471	11 %	3647	89 %	4118						471	11 %	3647	89 %	4118
	Female	798	17 %	3941	83 %	4740	798	17 %	3941	83 %	4740					
Age group	15-19	38	2 %	1656	98 %	1695	21	3 %	814	97 %	835	17	2 %	843	98 %	860
	20-29	241	9 %	2569	91 %	2809	158	11 %	1304	89 %	1461	83	6 %	1265	94 %	1348
	30-39	491	25 %	1482	75 %	1973	317	29 %	762	71 %	1079	174	19 %	719	81 %	893
	40-49	314	25 %	940	75 %	1254	188	28 %	488	72 %	675	127	22 %	452	78 %	579
	50+	185	16 %	942	84 %	1127	115	17 %	574	83 %	689	70	16 %	368	84 %	438
Household wealth quintile	Lowest	253	18 %	1148	82 %	1401	178	23 %	612	77 %	790	74	12 %	537	88 %	611
	Second	338	20 %	1349	80 %	1687	206	24 %	667	76 %	872	133	16 %	682	84 %	815
	Middle	323	18 %	1520	82 %	1843	182	20 %	730	80 %	911	141	15 %	791	85 %	932
	Fourth	268	13 %	1805	87 %	2073	170	15 %	969	85 %	1138	99	11 %	836	89 %	935
	Highest	88	5 %	1765	95 %	1853	63	6 %	964	94 %	1027	25	3 %	801	97 %	826
Educational attainment	No education	134	19 %	573	81 %	707	70	22 %	256	78 %	326	64	17 %	317	83 %	381
	Primary	426	19 %	1769	81 %	2196	266	23 %	866	77 %	1132	161	15 %	903	85 %	1064
	Secondary	664	13 %	4550	87 %	5214	438	15 %	2442	85 %	2879	227	10 %	2108	90 %	2334
	Higher	45	6 %	696	94 %	741	25	6 %	377	94 %	402	20	6 %	319	94 %	339
Marital status	Never married	532	10 %	4567	90 %	5099	341	14 %	2183	86 %	2524	191	7 %	2384	93 %	2575
	Currently married	507	17 %	2447	83 %	2955	259	16 %	1329	84 %	1587	249	18 %	1119	82 %	1367
	Formerly married	202	31 %	446	69 %	649	175	34 %	336	66 %	511	28	20 %	110	80 %	138
	Polygynous union	28	18 %	127	82 %	155	24	21 %	93	79 %	117	4	11 %	34	89 %	38
Number of trips during the last 12 months	No trips	823	15 %	4504	85 %	5327	563	19 %	2464	81 %	3027	260	11 %	2040	89 %	2300
	1-5	350	13 %	2242	87 %	2592	216	15 %	1242	85 %	1458	134	12 %	1000	88 %	1133
	6-10	38	9 %	370	91 %	408	12	8 %	143	92 %	155	26	10 %	228	90 %	254
	More than 10	56	11 %	450	89 %	506	6	7 %	90	93 %	97	50	12 %	359	88 %	409
	NA	3	13 %	22	87 %	25	1	27 %	2	73 %	3	2	11 %	20	89 %	22
Language group	Afrikaans	31	4 %	696	96 %	727	17	5 %	350	95 %	368	14	4 %	346	96 %	359
	Damara /Nama	86	9 %	913	91 %	999	52	9 %	502	91 %	555	34	8 %	411	92 %	445
	English	9	5 %	159	95 %	168	8	9 %	87	91 %	95	1	1 %	72	99 %	73
	Herero	42	7 %	596	93 %	638	28	8 %	329	92 %	357	14	5 %	267	95 %	281
	Kwangali	115	15 %	657	85 %	772	76	17 %	363	83 %	439	39	12 %	294	88 %	333
	Lozi	102	25 %	309	75 %	411	67	30 %	160	70 %	227	35	19 %	149	81 %	184
	Oshiwambo	824	18 %	3857	82 %	4680	517	21 %	1979	79 %	2496	307	14 %	1878	86 %	2185
	San	12	16 %	63	84 %	75	6	16 %	30	84 %	36	7	17 %	32	83 %	39
	NA	48	12 %	339	88 %	387	27	16 %	140	84 %	167	21	10 %	198	90 %	220

Lifetime number of sexual partners	No partners	19	2 %	1073	98 %	1092	12	2 %	514	98 %	526	7	1 %	559	99 %	566
	1-4	849	15 %	4730	85 %	5580	681	18 %	3099	82 %	3780	168	9 %	1631	91 %	1800
	5-20	274	17 %	1348	83 %	1623	76	22 %	265	78 %	342	198	15 %	1083	85 %	1281
	More than 20	45	23 %	147	77 %	193	3	22 %	11	78 %	14	42	24 %	137	76 %	179
	NA	82	22 %	289	78 %	371	26	33 %	52	67 %	79	56	19 %	236	81 %	293
Condom use	Does not use condom	789	14 %	4979	86 %	5768	518	16 %	2714	84 %	3232	271	11 %	2265	89 %	2536
	Uses condom	366	15 %	2035	85 %	2401	165	20 %	653	80 %	818	200	13 %	1382	87 %	1582
	NA	115	17 %	574	83 %	689	115	17 %	574	83 %	689					
Male circumcision	No	374	12 %	2672	88 %	3046						374	12 %	2672	88 %	3046
	Yes	96	9 %	959	91 %	1055						96	9 %	959	91 %	1055
	NA	800	17 %	3957	83 %	4757	798	17 %	3941	83 %	4740	2	10 %	16	90 %	18
Residence	Urban	613	15 %	3407	85 %	4020	409	19 %	1786	81 %	2195	204	11 %	1621	89 %	1824
	Rural	657	14 %	4181	86 %	4838	389	15 %	2155	85 %	2544	267	12 %	2026	88 %	2294
Distance to closest primary road	Less than 5 km	698	14 %	4292	86 %	4990	417	16 %	2222	84 %	2639	281	12 %	2070	88 %	2351
	5-50 km	545	15 %	3101	85 %	3646	366	18 %	1640	82 %	2005	180	11 %	1462	89 %	1641
	More than 50 km	26	12 %	196	88 %	222	15	16 %	80	84 %	95	11	9 %	116	91 %	127
Distance to closest city	Less than 5 km	252	14 %	1535	86 %	1787	157	16 %	843	84 %	1001	94	12 %	692	88 %	786
	5-150 km	916	15 %	5181	85 %	6097	578	18 %	2645	82 %	3224	337	12 %	2536	88 %	2873
	More than 150 km	102	11 %	872	89 %	974	63	12 %	453	88 %	515	40	9 %	419	91 %	459
Lifetime migration rate	Outflow more than 20%	150	16 %	803	84 %	954	104	19 %	438	81 %	543	46	11 %	365	89 %	411
	Outflow 1-20%	438	15 %	2417	85 %	2855	311	19 %	1309	81 %	1620	127	10 %	1108	90 %	1235
	Inflow 0-50%	324	13 %	2246	87 %	2570	195	15 %	1144	85 %	1339	130	11 %	1102	89 %	1231
	Inflow more than 50%	357	14 %	2123	86 %	2479	188	15 %	1050	85 %	1238	169	14 %	1072	86 %	1241
Administrative region	Caprivi	110	24 %	344	76 %	454	77	32 %	166	68 %	243	33	16 %	178	84 %	211
	Erongo	99	13 %	661	87 %	760	53	14 %	320	86 %	373	45	12 %	341	88 %	387
	Hardap	26	8 %	318	92 %	344	15	8 %	165	92 %	180	11	7 %	153	93 %	164
	Karas	41	12 %	303	88 %	345	26	14 %	156	86 %	181	16	10 %	148	90 %	163
	Khomas	242	12 %	1798	88 %	2040	127	12 %	906	88 %	1033	115	11 %	892	89 %	1007
	Kunene	23	9 %	224	91 %	247	12	9 %	125	91 %	137	11	10 %	99	90 %	110
	Ohangwena	130	16 %	689	84 %	819	102	21 %	386	79 %	488	28	9 %	302	91 %	331
	Okavango	124	17 %	627	83 %	752	79	18 %	354	82 %	433	45	14 %	274	86 %	319
	Omaheke	18	7 %	228	93 %	247	9	7 %	117	93 %	126	10	8 %	111	92 %	121
	Omusati	155	18 %	698	82 %	853	107	22 %	386	78 %	493	49	13 %	312	87 %	360
	Oshana	129	17 %	613	83 %	742	90	22 %	319	78 %	409	39	12 %	295	88 %	333
	Oshikoto	104	15 %	600	85 %	704	61	17 %	299	83 %	361	43	12 %	301	88 %	344
	Otjozondjupa	67	12 %	484	88 %	551	39	14 %	243	86 %	282	28	10 %	241	90 %	268

5.2.2 *Results of the logistic regression model*

In the following chapter, the results of the logistic regression analysis are presented. Three different models were compiled, one for all respondents, one for women and one for men. In table 4 are presented the odds ratios for the model parameters and their statistical significance according to p-value and 95 percent confidence interval for the odds ratios.

According to the results of the global model, women are 1.8 times more likely to be HIV positive than men. The results are significant at 99.9 percent confidence level. In regard to age group, respondents were divided to five groups. Among these groups respondents aged 30 to 39 are most likely to be HIV positive. This supports the tentative assumptions made according to descriptive statistics in the NDHS 2013 final report and based on frequencies in table 3. Compared to the reference group of respondents aged 15 to 19 years, the respondents aged 30 to 39 are more than nine times more likely to be HIV positive. For respondents aged 15 to 19 years the probability to be HIV positive is the smallest. In the male model, respondents aged 40 to 49 have highest HIV risk. When compared to the reference group respondents in this age group are more than five times more likely to be HIV positive. All results mentioned above are statistically significant with p-value smaller than 0.001.

According to the model, socioeconomic status has an effect on individual's HIV risk. The highest wealth quintile has evidently decreased HIV risk. The result is statistically significant at 9.99 percent confidence level. The reference group for wealth quintiles is the middle quintile. The lower wealth quintiles do not have statistically significant difference to the middle quintile in terms of HIV risk. Educational attainment has an effect on HIV risk in the model. The reference group for education level is respondents with secondary education as highest educational attainment, which is the most common level of education in the survey sample (table 3). Compared to the reference group, respondents whose highest educational attainment is primary education are 1.44 times more likely to be HIV positive than those with secondary education. The result is statistically significant with p-value smaller than 0.001. Moreover, respondents with education higher than secondary are significantly less likely to be HIV positive. The odds ratio for the parameter is 0.56 with p-value smaller than 0.05.

The language respondents speak as first language has some associations to HIV risk. Respondents who speak Oshiwambo, Lozi or San are most likely to be HIV positive. The reference group is respondents speaking Afrikaans. Among Afrikaans and English speakers are found the smallest percentage HIV positive in the survey sample (table 3). The results seem especially evident for women and for those who speak Lozi or Oshiwanbo. Women who speak Lozi are 3.8 times more likely to be HIV positive with p-value smaller than 0.01 and women who speak Oshiwambo 3.13 times more likely HIV positive with p-value smaller than 0.001. For men the associations are not as clear but respondents who speak San are more than four times as likely and those who speak Oshiwambo more than twice as likely to be HIV positive as reference group. Both ratios have p-value smaller than 0.05.

In regard to marital status, formerly married, which included divorced and widowed, are more likely to be HIV positive than respondents who are currently married. When women and men are examined separately the increased risk for formerly married can only be found in the female model. For formerly married women the risk to be HIV positive is more than twice as big as for those who are currently married. The odds ratio is statistically significant with p-value smaller than 0.001. Respondents currently in polygynous union were examined in the model separately. For this group no statistically significant results are found. Between currently married and never married respondents some differences in HIV risk level are found. For women those who were never married have slightly higher risk for HIV. This ratio was statistically significant with p-value smaller than 0.05. For men those who were never married have a decreased HIV risk in respect to those currently married with odds ratio of 0.65 with p-value smaller than 0.05.

Number of sexual partners during respondent's lifetime has a significant effect on their HIV risk in all the three models. The logic behind the effect of multiple sexual partners to HIV risk seems quite undisputed. The more sexual partners the respondent reported having, the higher is their risk for being HIV positive. Among men, respondents who reported having more than 20 sexual partners during their lifetime are almost three times more likely to be HIV positive than respondents who reported having had one to four partners during their lifetime. This odds ratio has a p-value smaller than 0.001. Also men who reported having 5 to 20 partners during their lifetime have higher risk for HIV. These respondents are 1.4 times more likely to be HIV positive than reference group. Among women the odds ratios are somewhat higher. Respondents with 5 to 20 partners are almost twice as likely to be HIV positive as the reference group. Respondents with more than 20 partners are more than four times as likely to be HIV positive as the reference group with p-value smaller than 0.05.

Even though in the global model parameters for all the groups in regard to lifetime number of sexual partners are statistically significant, in the female model the results are not as clear as in the male model. Also for women the odds ratios are higher but the parameters do not indicate as distinct differences between different groups and the odds ratios are not statistically significant. Nevertheless, in all three models, respondents who reported not having sex have significantly decreased risk for being HIV positive.

Use of condom in the last sexual intercourse is only included as an independent variable in the male model. This is because the question regarding condom use was not present in the data for over 50-year-old women. Due to missing data this variable was thus excluded from the global model and the female model. Nevertheless, even for men this variable does not produce statistically significant results but it would seem that respondents who used condom in their last sexual intercourse have somewhat higher odds ratio than reference group who did not use condom.

In the male model, also male circumcision is included as an independent variable. This variable did not produce statistically significant results but it would seem that male respondents who reported being circumcised were slightly less likely to be HIV positive. Nevertheless, the confidence interval for the odds ratio is quite wide and as stated before the results were not significant at 95 percent confidence level.

The interpolated estimates for HIV prevalence in different locations were used as an independent variable in the logistic regression model. Each survey cluster in the NDHS data were given an average HIV prevalence value of the 25-kilometre buffer area surrounding the cluster point. This HIV prevalence value would then be addressed to each respondent depending on which cluster they were interviewed in. The average HIV prevalence of the area has a statistically significant effect on individual's HIV risk in the logistic regression model. When the HIV prevalence in the area increases with one percentage point, the odds for the respondent to be HIV positive are 1.07 times higher. The odds ratios are similar for women and men separately. Parameters are statistically significant for global and female model at 99.99 percent confidence level and for male model at 99.9 percent confidence level.

Urban versus rural status of the survey cluster where respondent was interviewed does not produce statistically significant results in all the models. Nevertheless, the odds ratios would indicate that respondents living in rural areas have somewhat higher odds for being HIV positive. In the male model the odds ratio is smaller than in the global model and female model. Women living in rural survey clusters are 1.66 times more likely to be HIV positive than women living in urban clusters. Ratio is statistically significant at 99.9 percent confidence level.

Computational distance to nearest city according to OpenStreetMap data does not produce consistent results in the model. Odds ratios for all groups for this variable are close to one and not statistically significant. Either distance to primary road according to OpenStreetMap data does not produce statistically significant results in the model. Nevertheless, for this variable the odds ratios seem to be slightly more consistent. Proximity to primary road would seem to increase the risk for the respondent to be HIV positive. In other words, the further from the closest primary road is the survey cluster, the smaller is the risk for HIV. None of the results regarding computational distance from roads or cities were statistically significant.

Regional migration rate was available for the logistic regression model on constituency level and each survey cluster and thus respondent were given values depending on which constituency the survey cluster falls in. Migration rate was also observed earlier in this study. Migration rate indicates indirectly how much the population in the constituency has increased when number of people born in the constituency and number of people living there at the moment of the census count are compared.

Migration rate produces some statistically significant results in the model. Reference group for this variable is 1 to 20 percent outflow. Respondents who live in a constituency experiencing more than 50 percent inflow in migration rate are 1.39 times more likely to be HIV infected with respect to the reference group. The parameter for this group is statistically significant with p-value smaller than 0.05. When the reference group is compared with respondents who live in constituencies experiencing zero to 50 percent inflow, no significant differences are found. For those living in constituencies with more than 20 percent outflow, odds ratio seems to indicate higher risk for HIV. Nevertheless, this result is not statistically significant at 95 percent confidence level.

In all three models, respondents living in an area experiencing significant inflow have higher risk for HIV. This effect seems to be most evident in the male model. Men who live in areas experiencing more than 50 percent inflow are almost 2.3 times more likely to be HIV positive than men who live in areas experiencing minor outflow. Nevertheless, in the global model and for women, the results are not as consistent. In the global model, significant outflow in migration rate would seem to increase HIV risk more than minor inflow. For men, even minor inflow increases HIV risk more than outflow. Nevertheless, these differences are not statistically significant at 95 percent confidence level.

Table 4. Model parameters of the logistic regression model. Reference category in brackets.

	Global model		Female model		Male model	
	OR	95 % confidence interval	OR	95 % confidence interval	OR	95 % confidence interval
(Intercept)	0.00***	[0 - 0.01]	0.00***	[0 - 0.01]	0.01***	[0 - 0.03]
Sex (Male)						
Female	1.80***	[1.48 - 2.2]				
Age group (15-19)						
20-29	2.76***	[1.76 - 4.49]	3.44***	[1.94 - 6.57]	1.1	[0.57 - 2.27]
30-39	9.64***	[6.15 - 15.75]	12.32***	[6.92 - 23.6]	4.26***	[2.21 - 8.8]
40-49	8.79***	[5.47 - 14.64]	10.15***	[5.53 - 19.91]	5.11***	[2.55 - 10.89]
50+	3.95***	[2.39 - 6.76]	3.92***	[2.05 - 7.95]	3.83**	[1.78 - 8.69]
Wealth quintile (Middle)						
Lowest	0.93	[0.71 - 1.22]	1.17	[0.83 - 1.66]	0.74	[0.46 - 1.19]
Second	1.16	[0.92 - 1.45]	1.2	[0.88 - 1.63]	1.02	[0.71 - 1.46]
Fourt	0.83	[0.66 - 1.06]	0.85	[0.62 - 1.16]	0.68*	[0.46 - 0.99]
Highest	0.33***	[0.23 - 0.47]	0.43***	[0.28 - 0.67]	0.24***	[0.12 - 0.44]
Educational attainment (Secondary)						
No education	1.27	[0.95 - 1.7]	1.4	[0.91 - 2.14]	1.27	[0.8 - 1.98]
Primary	1.44***	[1.19 - 1.75]	1.6***	[1.23 - 2.07]	1.47*	[1.07 - 2.02]
Higher than secondary	0.56**	[0.35 - 0.85]	0.76	[0.43 - 1.26]	0.63	[0.3 - 1.23]
Marital status (Currently married)						
Never married	0.94	[0.77 - 1.14]	1.33*	[1.03 - 1.73]	0.65*	[0.46 - 0.91]
Formerly married	2.04***	[1.57 - 2.64]	2.75***	[1.99 - 3.8]	1.18	[0.64 - 2.1]
Currently in polygynous union	1.06	[0.63 - 1.75]	1.07	[0.56 - 1.96]	0.94	[0.15 - 3.69]
Number of trips during the last 12 months (No trips)						
1-5	0.78**	[0.65 - 0.94]	0.86	[0.68 - 1.08]	0.76	[0.54 - 1.05]
6-10	0.48**	[0.29 - 0.76]	0.44*	[0.19 - 0.92]	0.93	[0.49 - 1.67]
More than 10	0.71	[0.48 - 1.04]	0.19*	[0.03 - 0.67]	0.88	[0.55 - 1.37]
Language group (Afrikaans)						
Damara/Nama	1.32	[0.8 - 2.24]	1.33	[0.69 - 2.72]	1.25	[0.56 - 2.97]
English	2.72*	[1.11 - 6.16]	2.27	[0.69 - 6.51]	0	[0 - 0]
Herero	1.2	[0.68 - 2.16]	1.3	[0.61 - 2.81]	0.71	[0.22 - 2.14]
Kwangali	1.99*	[1.18 - 3.43]	2	[0.99 - 4.23]	1.78	[0.77 - 4.36]
Lozi	2.58**	[1.43 - 4.74]	3.8**	[1.73 - 8.67]	2.38	[0.9 - 6.52]
Oshiwambo	2.66***	[1.7 - 4.32]	3.13***	[1.71 - 6.13]	2.54*	[1.26 - 5.62]
San	2.91*	[1.14 - 6.98]	2.32	[0.55 - 8.05]	4.13*	[1.19 - 13.85]
Lifetime number of sexual partners (1-4)						
5-20	1.44**	[1.15 - 1.79]	1.95***	[1.34 - 2.83]	1.43*	[1.07 - 1.92]
More than 20	2.99***	[1.89 - 4.66]	4.35*	[0.92 - 16.04]	2.99***	[1.8 - 4.91]
No partners	0.44**	[0.23 - 0.77]	0.3**	[0.12 - 0.65]	0.4	[0.14 - 1]

Condom use in last intercourse (Did not use condom)						
Used condom					1.35	[1 - 1.82]
Male circumcision (Not circumcised)						
Circumcised					0.82	[0.57 - 1.17]
Residence (Urban)						
Rural	1.31	[0.98 - 1.76]	1.66**	[1.14 - 2.43]	1.02	[0.62 - 1.68]
Average HIV prevalence in a 25 km buffer area	1.07***	[1.05 - 1.1]	1.08***	[1.04 - 1.12]	1.06**	[1.02 - 1.11]
Distance to primary road (5-50 km)						
Less than 5 km	1.19	[0.93 - 1.52]	1.17	[0.85 - 1.61]	1.26	[0.81 - 1.96]
More than 50 km	0.84	[0.46 - 1.46]	0.92	[0.33 - 2.2]	0.42	[0.12 - 1.14]
Distance to a city (5-150 km)						
Less than 5 km	1.1	[0.84 - 1.43]	0.92	[0.65 - 1.31]	1.17	[0.74 - 1.82]
More than 150 km	1.18	[0.86 - 1.62]	1.26	[0.82 - 1.9]	1.03	[0.59 - 1.75]
Lifetime migration rate of the constituency where living (Outflow 1-20 %)						
Inflow 0-50 %	0.91	[0.72 - 1.14]	0.87	[0.65 - 1.16]	1.39	[0.94 - 2.08]
Inflow more than 50 %	1.39*	[1.06 - 1.81]	1.13	[0.79 - 1.62]	2.29***	[1.45 - 3.65]
Outflow more than 20 %	1.06	[0.81 - 1.39]	1.03	[0.73 - 1.43]	1.02	[0.59 - 1.72]
Predictive accuracy	85 %		79 %		88 %	
AIC	4119		2423.8		1512.6	

* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001

5.2.3 Assessment of model performance

For all three models presented in this study, several different candidate models with different selection of independent variables were tested. Goodness of fit for the candidates was tested using Akaike's Information Criterion and the statistical significance of the parameters for independent variables as well as accuracy of prediction of the models were assessed. Using these methods, the tested models were ranked and the final three presented in this study were chosen.

To estimate the accuracy of prediction 25 percent of the observations in the data were chosen with random selection and excluded from the computation of the model. This part of the data was used afterwards to test the predictive accuracy of the model. Overall, the accuracy of prediction gives relatively good results for the final models. The models predicted HIV status accurately for around 80 to 90 percent of the observations in the testing part of the data. According to the area under ROC curve the models did not perform as well but also this measure of model performance gives fairy good results ranging from 0.72 for male model to 0.77 for female model.

When different candidate models were tested, the different selections of independent variables did not produce major differences in the AIC value. For example, for the global model, the candidate model that had the highest AIC only had a statistical probability of 0.004 to minimise information loss in the same extent as the candidate that had the lowest AIC. In other words, the difference between these two models was high enough to consider dropping some variables. However, when the best performing model was compared with another candidate model with the second lowest AIC, the candidate had a probability of 0.4 to minimise information lost in the same extent as the first model. In other words, there is no statistical evidence that the first model would perform better (Hosmer et al. 2013).

Therefore, also the significance of the independent variables in the model and the accuracy of prediction for the different candidate models were assessed when choosing between the models which did not possess significantly different AIC values. In regard to accuracy of prediction, the different candidates performed also quite similarly. The difference of the observations predicted correctly was only few percentage points between candidate models. Hence, also some variables which did not produce statistically significant parameters were included in the model. This was done mainly to standardise the effect of these factors in model.

According to the accuracy of prediction, the male model was performing slightly better than the global model or female model respectively. Predictive accuracy was around 80 to 90 percent for all the models. Female model had lowest predictive accuracy. AIC values should only be compared when choosing from different candidates for the same model. Hence, differences in AIC cannot be used to compare the three models with each other.

According to area under Receiving Operating Characteristics (ROC) curve, the female model got the best results from the three models (figure 29). It should be noted, that predictive accuracy and AUC value measure model performance by different criteria and can give different results. Also comparison between different models according to AUC value is not ideal. Nevertheless, it is good to acknowledge that with different criterion the performance for the models can have different results (Hosmer et al. 2013).

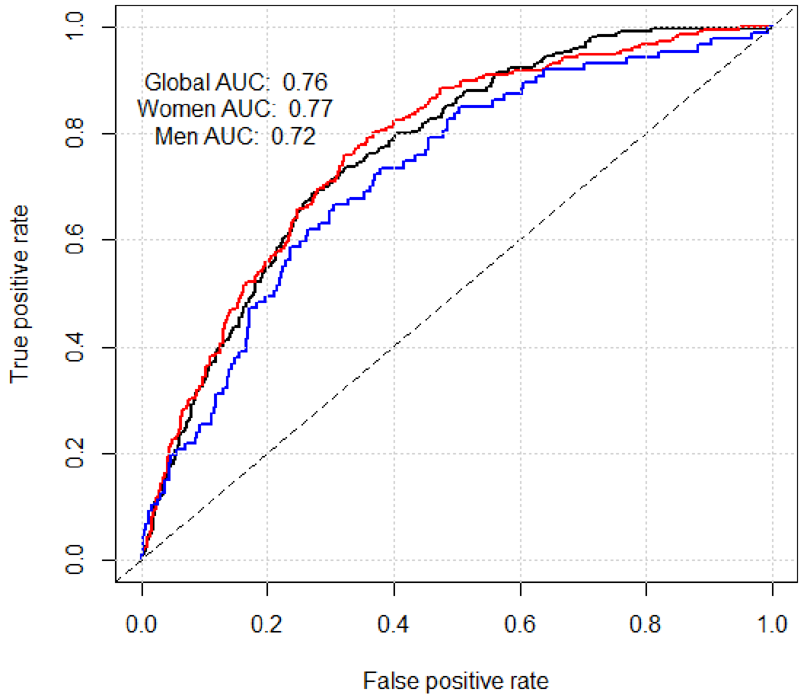

Figure 29. ROC curves and AUC values for the three logistic regression models.

6 Conclusions and discussion

In this study, I have studied the characteristics and spatial distribution of the HIV epidemic in Namibia. The objective of the study was to examine the spatial variation of HIV prevalence with new methods that had not been utilised before in Namibia. This is mainly because previously nationally representative data with both HIV testing and georeference has not been available. The spatial approaches of this study focused on examining differences in geographical distribution of HIV prevalence and producing interpolated sub-regional estimates. Other key objective was to recognise sociodemographic and geographical factors that affect HIV risk for an individual person in Namibia. These examinations aimed to produce information about which factors and processes currently drive the diffusion of the epidemic.

In this final part of this study, I will present the key findings of this study. The main focus is on how the results supplement existing knowledge of the HIV epidemic in Namibia and elsewhere in developing countries of the Global South. The results of the logistic regression analysis are compared with the results from the final report of the NDHS 2013 as well as from previous study results. Visual interpretation of the spatial variation of the HIV prevalence is assessed and compared to how the independent variables used in the logistic regression model vary geographically according to the NHS 2013 final report (MoHSS and ICF International 2014) and the 2011 census (NSA 2013). The objective behind these visual examinations was to make tentative interpretations of how these variables included in the regression model affect the epidemic on spatial scale. In other words, to answer the question, why the interpolated sub-regional estimates for HIV prevalence variate geographically the way they do. This is not a valid method for making exact conclusions but for the purposes of this study it provided an overview of the nature of HIV epidemic in Namibia.

For the spatial examinations, the study question was whether there exists specific locations of high or low HIV prevalence in comparison to the surrounding administrative region. The key point of interest was to find out whether these locations differ substantially from the average HIV prevalence of the administrative region. The answer to this study question would add on the current knowledge on the HIV epidemic in Namibia. Currently the spatial variation of the epidemic is only known on administrative region scale.

Interpolated estimates unveil to some extent the spatial variation in HIV prevalence inside the national borders of Namibia and inside the administrative regions. This gives a more detailed picture of the spatial nature of the epidemic in Namibia. According to the results, the interpolated sub-regional HIV prevalence varies substantially inside the administrative areas. Especially in North-Central Namibia where the average HIV prevalence for the administrative regions is high compared to other regions, the interpolated estimates for HIV prevalence vary greatly. Also inside regions where average HIV prevalence is relatively low, locations with high HIV prevalence according to interpolated estimates can be found.

Outside the North-Central Namibia and the Caprivi strip, which hold the highest average HIV prevalence, clusters with high HIV prevalence concentrate in the proximity of big cities. According to the interpolated estimations for HIV prevalence, the computed values are highest where there is high population density or urban centre present. In fact, according to visual interpretation, in North-Central Namibia interpolated estimates for HIV prevalence depend more on proximity of urban centre than high population density alone. Yet this was not unambiguous. In other parts of Namibia, some urban centres with high population density and growth rate experience relatively low HIV prevalence. According to visual interpretation, it was not possible to recognise whether proximity of urban centres had an effect on HIV prevalence.

According to previous study results, urban areas in some cases foster higher HIV prevalence than rural areas (Arroyo et al. 2005). Surprisingly, in Namibia the average HIV prevalence was higher in rural than in urban survey clusters according to the results of the NDHS 2013 final report (MoHSS and ICF International 2014). For women, HIV prevalence in rural areas was higher than in urban areas. For men, the opposite results were found. The HIV prevalence was higher for men in urban areas than rural areas. This needs to be taken into account when differences regarding HIV prevalence in urban and rural areas are examined.

The examinations of sub-regional interpolated estimates for HIV prevalence show that urban versus rural place of living did not have as unambiguous effect on HIV prevalence as previous study results have shown. When this phenomenon was examined through the sub-regional estimates for HIV prevalence, it seemed that in the proximity of urban concentrations HIV prevalence was higher. According to visual interpretation of HIV prevalence on several different scales of observation urban areas would seem to foster higher estimates for HIV prevalence. This is in conflict with the results from the NDHS 2013 final report. In the logistic regression model urban versus rural place of living produced inconsistent results that were not statistically significant. Nevertheless, according to the model, urban versus rural place of living seemed to affect HIV risk differently for women and men. This would indicate that this factor affects the epidemic in a more complex manner than previously thought.

According to more in-depth examinations of the interpolated sub-regional HIV prevalence, there seems to exist variation inside urban concentrations. For example in Windhoek, there are significant differences in the level of HIV prevalence between the southern and northern parts of the city. Northern parts accommodate significantly higher HIV prevalence than the southern parts of the city. Most likely the southern parts of the city are upper class residential areas.

These findings could be resulting from scattered population distribution. Urban areas are usually surrounded with higher population densities. The urban areas could be high lightened in visual interpretations only for this reason. It is possible that actually rural areas with high population density right outside urban concentrations foster the higher HIV prevalence instead of the survey clusters classified as urban.

In addition, North-Central Namibia and Caprivi, which hold the highest values for HIV prevalence, are mostly classified as rural areas. When the sub-regional estimates for HIV prevalence are examined, areas with high HIV prevalence in North-Central Namibia and Caprivi also extend to areas outside urban centres even though also in these areas proximity to urban centre would seem to increase HIV prevalence.

These results also raise the question, why in Namibia conflicting results are found in respect to how urban versus rural place of living affects the HIV epidemic, when in other developing countries opposite results have been found (Bärnighausen et al. 2007; Niragire et al. 2015). Could the reason behind differing results originate from differences in what kind of areas are classified as urban or rural? These indistinct connections between high HIV prevalence and urban versus rural environment would need further examinations.

Population mobility has traditionally been studied extensively as a contributor to the HIV epidemic. Nevertheless, the ways in which migration patterns and tendency for individuals to move from place to place affects HIV risk is a complex phenomenon to grasp. High migration rates are usually associated with urban areas rather than rural areas. High migration rates and access to means of transportation in the form of proximity to primary road network, have been found to correlate with high HIV prevalence (Tanser et al. 2000; Arroyo et al. 2006; Bärnighausen (2007); Feldacker et al. 2010).

In Namibia, the main migration flows exist between North-Central Namibia and Khomas region where the capital Windhoek is located (figure 11). Through these routes have in the long term migrated the largest number of individuals. The number of people have been estimated according to how many people were born in specific region and how many of those people currently live in another specific region. These figures represent absolute numbers of people moving in long term.

In the logistic regression model, relative migration rates have been used instead of absolute number of people moving. These implied migration rates represent the number of people moving to the area in respect to 100 people born in the area. According to migration rates for long-term population growth, North-Central Namibia seems to exhibit most outflow and especially Khomas region where Windhoek is located experiences most inflow.

According to the NDHS 2013 final report, the respondent's tendency to travel and spend time away from home did not correlate to HIV risk as would be expected. Among those who reported having made a lot of trips and spending nights away from home the average HIV prevalence was actually lower. This is in conflict with earlier study findings from other developing countries in the Global South (Coffee et al. 2005, 2007; Bärnighausen et al. 2007).

In the logistic regression model, the effect of migration patterns and population mobility on the HIV epidemic was estimated according to the number of trips respondent had done during the last 12 months and the regional migration rate. As the final report of the NDHS 2013 indicated, increasing number of trips during the last 12 month actually decreased the risk for the respondent to be HIV positive. However, the results were somewhat inconsistent.

Regional migration rates on the other hand seemed to affect the HIV epidemic more clearly. Especially areas experiencing significant inflow seemed to foster higher HIV prevalence. Nevertheless, also with this factor major differences between men and women were found. For men, it seemed that areas experiencing inflow caused higher risk for HIV when for women areas experiencing significant outflow also caused escalated risk for HIV infection. This phenomenon could originate from the fact that for men urban areas foster higher HIV prevalence and for women the rural areas.

Access to means of transportation has been traditionally studied by examining proximity to primary road network. The distance to nearest primary road has been used to estimate HIV risk in several studies (Tanser et al. 2000; Arroyo et al. 2006; Feldacker et al. 2010). This approach has been especially popular when spatial methods have been pioneered in the studies of the HIV epidemic. According to previous study results proximity to primary roads did increase HIV risk but the results were not coherent in all the studies (Arroyo et al. 2006; Feldacker et al. 2010).

Also in this study, proximity to primary roads has been used to predict HIV risk. According to visual interpretation, connections between occurrences of high HIV prevalence and proximity of primary road are difficult to interpret. This is mainly because the population in Namibia is highly concentrated to specific areas. These areas of dense population also foster high HIV prevalence values and usually primary road lies near. Also according to the logistic regression model, connections between proximity to primary road and increased HIV risk could not be verified. The model parameters indicate that respondents interviewed in survey clusters closer to primary road would be more likely to be HIV infected but the results were not statistically significant.

Demographic factors have been studied a lot in regard to HIV epidemic and HIV risk for individuals. Multiple studies have shown that in developing countries of the Global South women have higher risk for HIV infection than men do (Hargreaves et al. 2002; Arroyo et al. 2005; Montana et al. 2007; Messina et al. 2010; Aulagnier et al. 2011; Magadi and Desta 2011). In regard to age, it has been detected that ages from approximately 30 to 40 are most likely to be HIV positive in the countries of SSA. For men the group with highest HIV prevalence are slightly higher (Magadi and Desta 2011). The results of this study support these earlier findings.

Regional population structure in Namibia shows that in North-Central Namibia, number of women often exceeds that of men (NSA 2013). Also in these areas, the general HIV prevalence seems to be higher than elsewhere in Namibia. According to the separately interpolated HIV prevalence for women and men, the higher proportion of women seems to raise the HIV prevalence in northern Namibia. It is also interesting that the HIV prevalence around the city of Rundu is higher than in other large urban centres in Namibia. In this area proportion of women is higher than in other urban centres. Rundu is the second largest city in Namibia. In other words, population structure in regard to sex seems to affect the dynamics of the HIV epidemic.

It is also no surprise that according to the logistic regression model, women in Namibia have higher HIV risk than men. These results support earlier study results from other developing countries. As was stated before, almost without exception developing countries where HIV risk has been studied using DHS data women had higher probability for being HIV positive than men. This trend also exist in Namibia according to NDHS 2013 final report and the results of this study.

The median age of communities in North-Central Namibia is younger than elsewhere in Namibia. However, when the age structure for different regions in Namibia is examined more in-depth, regions with large urban centres such as Erongo or Khomas have a high proportion of young adults and regions with low median age actually foster a large proportion of children. In regions of North-Central Namibia, live also relatively many people aged 80 years or more than other regions outside this area.

Even though HIV prevalence is highest in the regions of North-Central Namibia and Caprivi, in these regions a relatively small proportion of population are from the age groups which hold the highest HIV risk according to the NDHS 2013 final report and the logistic regression analysis of this study. In fact, more individuals from these ages actually reside in urban areas (NSA 2013). Thus when examined regionally no clear pattern can be found between median age of the community or size of specific age groups and the level of HIV prevalence.

One common theme that has been found to influence HIV risk according to previous studies is socioeconomic characteristics. Demographics has also been found to have an effect on how the socioeconomic background affects individual’s HIV risk (Messina et al. 2010; Magadi and Desta 2011). Income level affects HIV risk in a complex manner and in fact socioeconomic background seems to affect HIV risk indirectly through altered sexual behaviour (Hargreaves et al. 2002). Most studies have found that individuals with higher income level tend to have higher risk to be HIV positive (Shelton et al. 2005; Mishra et al. 2007). This has also been found on national scale. Countries of the SSA with higher GNI per capita had higher HIV prevalence (Fox 2010).

In Namibia, according to the NDHS 2013 data, the level of income is significantly higher in urban areas than in rural areas. When income level is examined regionally, in North-Central Namibia fewer individuals are in the highest wealth quintile and respectively in urban Khomas and Erongo regions as many as half of the population are in the highest wealth quintile. Indeed, in Namibia it would look as if on regional scale the wealthier areas would foster lower HIV prevalence values. When the level of educational attainment is examined regionally, it seems that the same areas that accommodate high income level also accommodate high level of educational attainment.

According to the logistic regression model, these connections between socioeconomic status and HIV risk also exist on individual person scale. The results of the model show that the risk for HIV is undoubtedly lowest in the highest wealth quintile. In regard to educational attainment, similar results are found. The higher the educational attainment the lower the risk for the respondent to be HIV infected. These factors were among the few that had clear effect on individual's HIV risk and produced highly significant results in the logistic regression model. Also, when the regional variation in the socioeconomic factors and the results of the logistic regression model are compared with each other, the findings seem to support each other. In other words, in Namibia socioeconomic background would seem to play a major role in the dynamics of the HIV epidemic.

In addition to income level and level of educational attainment, marital status is a common characteristic of an individual that has been used in many studies to predict HIV risk. This factor also explains indirectly individual's sexual behaviour. According to previous study results, divorced or especially widowed individuals had in some cases higher HIV risk than their counterparts did (for example Magadi and Desta 2011; Mmbaga 2013; Tenkorang 2014; Kim et al. 2016). Results from Namibia support these earlier findings. According to the final report of the NDHS 2013, HIV prevalence is significantly higher among those who are divorced or especially widowed. The results of the logistic regression model are in unison with the reported HIV prevalence in NDHS 2013 final report even though the results were statistically significant only in the female model.

According to the NDHS 2013 final report, the group of respondents who were widowed had higher HIV prevalence than those divorced. This indicates that the increased risk for HIV for those who were formerly married would be due to the effect of widowed respondents in this group. In some cases it could be that the spouses of these respondents actually died of AIDS but this was not further investigated. Also living in polygynous union has been in some cases noted to increase HIV risk (Reniers and Tfaily 2012) even though also contrary findings exist (Maher et al. 2011; Sawers and Isaac 2017). Also in Namibia, according to the final report of the NDHS 2013, HIV prevalence seems to be slightly higher among survey respondents who were living in a polygynous union. However, the logistic regression model of this study did not support this assumption.

As was also stated in the theory background of this study, many of the demographic and socioeconomic factors are most likely causing increased HIV risk indirectly. Many previous studies regarding HIV risk have proposed that these factors most likely affect sexual behaviour of the individuals and through altered behaviour indirectly increase the risk for HIV infection (Hargreaves et al. 2002; Arroyo et al. 2005; Feldacker et al. 2010). Non-marital sexual relations (Johnson et al. 2017), number of concurrent or lifetime sexual partners (Kim et al. 2016) and age at initiation of sexual life (Hossain 2014; Niragire et al. 2015) have been found in some cases to correlate with increased HIV risk. Nevertheless, sexual behaviour and networks are a difficult phenomenon to study and the results have been conflicting which speaks for the complexity of the phenomenon in question.

In this study, sexual behaviour was not examined on regional scale due to lack of adequate data. However, factors indicating difference in sexual behaviour of individual respondents were included in the logistic regression model. In Namibia, increasing number of sexual partners during respondent's lifetime increased the probability for them to be HIV positive. This connection seems consistent and undisputed and it is in line with earlier study results. Condom use was only included in the male model where it did not produce statistically significant results, but it seemed that respondents reporting to have used condom in their last sexual intercourse had higher odds to be HIV positive. Controversial results like this have been found also in earlier studies which have predicted HIV risk with condom use (Kim et al. 2007; Johnson et al. 2017). In fact, condom use is usually included only in studies concerning high-risk or HIV positive populations because of its troublesome nature as predictive variable for HIV risk (Boerma and Weir 2005).

In addition to these factors, also other variables indicating sexual behaviour were tried in the candidate models for the regression analysis. These included number of sexual partners during the last 12 months and age at first sexual intercourse. These factors did not produce consistent or statistically significant results and seemed to reduce the reliability of the model. According to the NDHS 2013 final report, among survey respondents who had their first sexual intercourse before the age of 15 the HIV prevalence was lower. This is in conflict with earlier study results which have proposed that starting sex life at young age would increase the odds of being HIV positive (for example Niragire et al. 2015).

Sexual behaviour and networks is a complex study subject and this study failed to open this phenomenon up further. Nevertheless, connections between HIV risk and sexual behaviour were not in clear conflict with results found in other developing countries. It should be noted, that in the logistic regression model of this study many variables such as computational distance from road network or major cities among others did not produce statistically significant results in the model. Hence, it might be that sexual behaviour of individuals plays a bigger role among the factors affecting the HIV epidemic than geographical location and access to means of transportation.

The underlying mechanisms through which socioeconomic, demographical and geographical factors affect the HIV epidemic have been assessed in this study through individual survey respondent scale logistic regression model. To some extent the spatial nature of the HIV epidemic has been included in the logistic regression model as a variable that indicates the average level of HIV prevalence in a 25-kilometre buffer area surrounding each survey cluster and hence respondent. High HIV prevalence level in the surroundings of the survey clusters increase respondent’s odds to be HIV positive. This approach has been used also in other studies using DHS data to study HIV epidemic and the results have been similar (for example Messina et al. 2010). HIV prevalence in NDHS survey clusters was also tested for spatial autocorrelation and the results indicated correlation between HIV prevalence and its spatially lagged counterpart in survey cluster points (figure 25).

The motivation behind this study arose from the publishing of the NDHS 2013 data in the end of 2014. This new data provided for the first time in Namibia spatially referenced nationally representative survey data that included results from HIV testing. In Finland, there exists a long research tradition on the population dynamics and history in Namibia. HIV epidemic in Namibia has been studied before as a part of this long research history. When this new data was published there came to be a demand for examining the characteristics and spatial dynamics of the HIV epidemic in Namibia with more in-depth.

The Demographic and Health Survey data has been used a lot in studying of the HIV epidemic in developing countries of the Global South and SSA especially. There exists a lot of previous research that has used the data with various methods. From these previously used methods were selected the ones used in this study. The method for interpolating the HIV data of the NDHS 2013 has been designed especially for the DHS data (Larmarange et al. 2011). Logistic regression is a common statistical method used in demographical studies. It is also the most common method used for the DHS data and it has been applied also in HIV studies.

Nevertheless, during recent years HIV studies using more sophisticated spatial methods have been published. These methods include for example spatial regression and they have been applied also to DHS data after the georeference of these data has become more available and reliable. In these spatial applications of regression a spatially lagged dependent variable is included on the right side of the regression equation (Chimoyi and Musenge 2014; Barankanira et al. 2016). This way the model takes into account spatial auto-correlation of the dependent variable. These new methods include spatial applications of general linear regression as well as logistic regression. For this study, some of these methods were tested but the final selection of methods was restricted to more basic statistical models. For further research these methods should be definitely considered also in Namibia.

I decided to use the classical logistic regression on individual person scale. The main reason behind this choice was that the NDHS 2013 data was georeferenced on survey cluster scale and all the information available was in these survey clusters. It proved to be problematic to find adequate spatially georeferenced data where the information would be on different spatial scale compared to the NDHS data. Also the scattered population in Namibia made it difficult to apply spatial regression when regression was applied to standard size regional units. When regression was applied to survey clusters problem arose from the quantity of clusters with very few survey respondents. Nevertheless, spatially lagged logistic regression model could have worked for this study and the next step could be to apply this method to studying the HIV epidemic in Namibia.

The methods chosen for this study can be seen as safe bets. For the other possible applications that were tested, several kind of problems occurred and the model performance did not seem to be adequate. With the classical logistic regression, better model performance was achieved and using the interpolated HIV prevalence as an independent variable in the model took into account the spatial nature of the dependent variable and the level of HIV prevalence in the surroundings of the survey cluster. In other words, the odds ratios for the model parameters present the effect different variables have on the HIV risk regardless of the HIV prevalence in their surroundings.

For the logistic regression model there also exist some vulnerabilities because HIV epidemic is a complex phenomenon to model with statistical methods. Nevertheless, the model was constructed carefully and as many variables were included as was available in the DHS data keeping in mind to avoid internal correlation. For example population density was considered for the regression model but it was found to correlate with urban versus rural place of living. Population density was on different spatial scale as the NDHS data and did not seem as reliable. Therefore it was excluded from the analysis.

However, other limitations still exist for this study. These include firstly the vulnerabilities of the data used which were presented more in-depth in the data and methods part of this study. For the classical logistic regression model the DHS data works quite well when it is used in its original form on individual person scale. Nevertheless, sources of error have been found for the DHS data. These are mainly problems that occur when data from different countries and between different years are compared due to differences in practices. For example laboratory error which is resulted from differences in settings and practices falls among these.

However, sources of error that would have most effect on the results of this study include for example non-response bias especially in the case of HIV testing. In some countries where HIV testing have been done as a part of the DHS studies also evaluations of non-response bias were published. In Namibia such evaluation has not yet been available.

Previous studies have examined the effects of the non-response bias to HIV prevalence in DHS data and usually the error margins have been found to be marginal. Also no major differences in socioeconomic status were found between different survey sample populations and non-respondents (Ministry of Health, Kenya 2005; Marston et al. 2008).

In this study, especially sources of error that emerge when spatial data is used need to be taken into account. The first problem with georeference occurred within the NDHS 2013 data itself. The survey clusters are given georeference according to their location but this location does not represent the actual location of where the respondent is living. Furthermore, this location is manipulated in order to protect the privacy of the respondents. The manipulated location of the survey cluster is put around the actual location somewhere inside a sphere of few kilometres.

Source of error also arose from combining the DHS data with spatial data from other sources. The primary road network and exact locations of cities and towns were downloaded from the OpenStreetMap project. When this data was combined with the survey data each cluster was given a distance from the nearest primary road. For reasons explained earlier this can cause error since the spatial reference of the survey clusters is not perfect.

The main value of this study was to reach a more in-depth picture of the HIV epidemic in Namibia. The key achievement was to produce interpolated sub-regional estimates of the HIV prevalence in Namibia according to the new NDHS 2013 data. Previously adequate data for producing corresponding analysis has not been available. These sub-regional estimates now provide more in-depth picture of the spatial variation of occurrences of high and low HIV prevalence in Namibia instead of statistics published on administrative region scale. Especially differences in HIV prevalence for urban and rural areas can be assessed more effectively though these new estimates. Indeed, the most interesting finding that could be made through the results of this study was the manifold connections between HIV prevalence and urban versus rural place of living.

Despite their vulnerabilities, HIV testing modules in DHS studies are currently one of the most reliable sources for estimating regional HIV prevalence in developing countries of the Global South. Where HIV testing has been done as a part of the DHS studies, logistic regression models have been used to predict the effect sociodemographic factors have on

HIV risk for individuals. In Namibia this had not yet been done. The logistic regression analysis conducted in this study gives similar results as corresponding studies conducted elsewhere in developing countries of the Global South. To a great extent, the connections between individual scale socioeconomic and demographic characteristics as well as geographical characteristics of the surroundings affect individual's HIV risk in a similar way in Namibia as in corresponding studies from elsewhere.

The results of this study reveal new information on the structures and means of effect behind the dynamics of the HIV epidemic in Namibia. When demographics and socio-economic features of different areas are known this information can also help in focusing preventative actions in these areas and among correct demographic groups. For example new innovations such as the Pre-exposure prophylaxis (PrEP) medicine can be recommended among population groups and in areas where the risk for HIV has been found to be highest (Cremin et al. 2017).

According to UNAIDS (2017a) new HIV positive cases recognised in Namibia are decreasing. Also number of people dying of AIDS is decreasing significantly in Namibia as well as globally. Nevertheless, the HIV prevalence levels in developing countries of the Global South and especially in SSA are still high. In Namibia, the level of HIV prevalence is nationally lower than in neighboring countries but still high when compared to the average of the continent. It is also important to recognise the differences that exist inside the national boundaries and also inside administrative regions. Also this information can help the government target preventative measures more effectively.

In this study, I have utilised open source software and data. The software includes for example R statistics, QuantumGIS and GeoDa. Open source data have been acquired from GADM and OpenStreetProject. Open source software and data are becoming increasingly common in several fields from commercial to academic sector. In this study, I wanted to demonstrate that with the DHS data available from the program, open source software and additional open source data, spatial approach on the HIV epidemic is possible even without vast resources. Especially in developing countries utilisation of open source opens up possibilities that would otherwise be limited.

The results of this study to a large extent confirm findings from other developing countries of the Global South even though characteristics unique for the epidemic in

Namibia were also found. The dynamics of the HIV epidemic are difficult to grasp and it is important that even though the situation is improving these complexities are challenged and the phenomenon further investigated. New study results are needed constantly to give information of the changing situation and provide up to date knowledge for decision making in public health care and improvement strategies. Especially in Namibia, where nationally representative survey data with HIV testing was published for the first time, it is important to keep making effort in understanding what drives the epidemic and what are the best ways to make intervention.

References

Anselin, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical analysis, Vol 27, No. 2* (April 1995) Ohio State University Press.

Arroyo, M., Hoelscher, M., Sateren, W., Samky, E., Maboko, L., Hoffmann, O., Kijak, G., Robb, M., Birx, D. and McCutchan, F. (2005). HIV-1 diversity and prevalence differ between urban and rural areas in the Mbeya region of Tanzania, *Aids, vol. 19, no. 14, pp. 1517-1524.*

Arroyo, M.A., Sateren, W.B., Serwadda, D., Gray, R.H., Wawer, M.J., Sewankambo, N.K., Kiwanuka, N., Kigozi, G., Wabwire-Mangen, F., Eller, M., Eller, L.A., Birx, D.L., Robb, M.L. and McCutchan, F.E. (2006). Higher HIV-1 incidence and genetic complexity along main roads in Rakai District, Uganda, *Jaids-Journal of Acquired Immune Deficiency Syndromes, vol. 43, no. 4, pp. 440-445.*

Atashili J. (2006). Adult male circumcision to prevent HIV? *International Journal of Infectious Diseases, 10 (3) (2006), pp. 202-205.*

Aulagnier M, Janssens W, De Beer I, van Rooy G, Gaeb E, et al. (2011). Incidence of HIV in Windhoek, Namibia: Demographic and Socio-Economic Associations. *PLoS ONE 6(10): e25860.*

Barankanira E., Molinari N., Niyongabo T. and C. Laurent (2016). Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention. *BMC Public Health. 2016; 16: 118*. Published online 2016 Feb 4.

Bauer G. (1998). *Labor and democracy in Namibia 1971-1996*. 239 p. Ohio University Press.

Boerma S. and S. Weir (2005). Integrating demographic and epidemiological approaches to research on HIV/AIDS: the proximate determinants framework. *The Journal of Infectious Diseases, 191 (Suppl. 1) (2005), pp. S61-7.*

Busgeeth, K. and U. Rivett (2004). Research: The use of a spatial information system in the management of HIV/AIDS in South Africa. *International Journal of Health Geographics 2004, 3:13.*

Buvé A, Bishikwabo-Nsarhaza K and Mutangadura G (2002). The spread and effect of HIV-1 infection in sub-Saharan Africa. *Lancet. 2002 Jun 8; 359(9322):2011-7.*

Bärnighausen, T., Hosegood, V., Timaeus, I. M., and Newell, M.-L. (2007). The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *AIDS (London, England), 21(Suppl 7), S29–S38.*

Chambless, L. E. and Boyle, K. E. (1985): Maximum Likelihood Methods for Complex Sample Data: Logistic Regression and Discrete Proportional Hazards Models. *Communications in Statistics - Theory and Methods 14, 1377-1392.*

Chimoyi D.F. and Musenge E. (2014). Spatial analysis of factors associated with HIV infection among young people in Uganda, 2011. *BMC Public Health. 2014;14(1):555.*

Coffee, M., Garnett, G., Mlilo, M., Voeten, H., Chandiwana, S. and Gregson, S. (2005). Patterns of movement and risk of HIV infection in rural Zimbabwe. *Journal of Infectious Diseases, vol. 191, pp. S159-S167.*

Coffee, M., Lurie, M.N. and Garnett, G.P. (2007). Modelling the impact of migration on the HIV epidemic in South Africa. *Aids, vol. 21, no. 3, pp. 343-350.*

Cohen M.S., Chen Y.Q., McCauley M., Gamble T., Hosseinipour M.C., Kumarasamy N., Hakim J.G. et al. (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *New English Journal of Medicine. 2011 Aug 11; 365(6):493-505.*

Cremin I., McKinnon L., Kimani J., Cherutich P., Gakii G., Muriuki F., Kripke K., Hecht R., Kiragu M. and J. Smith (2017). PrEP for key populations in combination HIV prevention in Nairobi: a mathematical modelling study. *LANCET HIV Volume: 4 Issue: 5 Pages: E214-E222.*

Cuadros D.F., Awad S.F. and L.J. Abu-Raddad (2013). Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa. *Int J Health Geogr. 2013;12(1):28.*

Cuadros, D. F., Li, J., Branscum, A. J., Akullian, A., Jia, P., Mziray, E. N., and Tanser, F. (2017). Mapping the spatial variability of HIV infection in Sub-Saharan Africa: Effective information for localized HIV prevention and control. *Scientific Reports, 7, 9093.*

De Walque, D. (2002). How Does the Impact of an HIV/AIDS Information Campaign Vary with Educational Attainment? Evidence from Rural Uganda. *Journal of Development Economics Volume 84, Issue 2, November 2007, Pages 686-714.*

Deheneffe, J.-C., Carael, M. and A. Noumbissi, (1998*). Socioeconomic Determinants of Sexual Behaviour and Condom Use*. World Bank, Washington, D.C., USA.

Feldacker, C., Emch, M. and Ennett, S. (2010). The who and where of HIV in rural Malawi: Exploring the effects of person and place on individual HIV status. *Health & Place, vol. 16, pp. 996-1006.*

Finckenstein, B. (1990). AIDS in Namibia, in Stein, Z. and Zwi, A.. *Action on AIDS in Southern Africa, pp. 46-8.*

Fox, A.M. (2010). The Social Determinants of HIV Serostatus in Sub-Saharan Africa: An Inverse Relationship Between Poverty and HIV? *Public health reports, vol. 125, pp. 16-24.*

Frayne, B. (2004). Migration and urban survival strategies in Windhoek, Namibia. *Geoforum 35 (2004) pp. 489-505*.

Fylkesnes K., Musonda R.M., Sichone M., Ndhlovu Z., Tembo F. and M. Monze (2001). Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys. *AIDS. 2001 May 4;15(7):907-16.*

Fylkesnes K., Ndhlovu Z., Kasumba K., Mubanga Musonda R. and M. Sichone (1998). Studying dynamics of the HIV epidemic: population-based data compared with sentinel surveillance in Zambia. *AIDS. 1998 Jul 9;12(10):1227-34*.

Garciá-Calleja J. M., Gouws, e. and Ghys, P. D. (2006). National population based HIV prevalence surveys in sub-Saharan Africa: results and implications for HIV and AIDS estimates. *Sexually Transmitted Infections 2006;82(Suppl III):iii64-iii70.*

GeoDa (2017). <*https://geodacenter.github.io/*>. Retrieved. 20.10.2017.

Gould, W. and Woods, R. (2003). Population geography and HIV/AIDS: The challenge of a 'Wholly exceptional disease'. *Scottish Geographical Journal, vol. 119, no. 3, pp. 265-281.*

Gray RH1, Wawer MJ, Serwadda D, Sewankambo N, Li C, Wabwire-Mangen F, Paxton L, Kiwanuka N, Kigozi G, Konde-Lule J, Quinn TC, Gaydos CA and McNairn D. (1998). Population-based study of fertility in women with HIV-1 infection in Uganda. *Lancet. 1998 Jan 10;351(9096):98-103.*

Gray, R.R., Tatem, A.J., Lamers, S., Hou, W., Laeyendecker, O., Serwadda, D., Sewankambo, N., Gray, R.H., Wawer, M., Quinn, T.C., Goodenow, M.M. and Salemi, M. (2009). Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *Aids, vol. 23, no. 14, pp. F9-F17.*

Hargreaves, J., Morison, L., Chege, J., Rutenburg, N., Kahindo, M., Weiss, H., Hayes, R., Buve, A. and Study Grp Heterogeneity HIV Epidem (2002). Socioeconomic status and risk of HIV infection in an urban population in Kenya. *Tropical Medicine & International Health, vol. 7, no. 9, pp. 793-802.*

Harling, G., Newell, M., Tanser, F., Kawachi, I., Subramanian, S.V. and Baernighausen, T. (2014). Do Age-Disparate Relationships Drive HIV Incidence in Young Women? Evidence from a Population Cohort in Rural KwaZulu-Natal, South Africa. *Jaids-Journal of Acquired Immune Deficiency Syndromes, vol. 66, no. 4, pp. 443-451.*

Hellmuth, M. E. (2000). Water resources of Namibia. In: *Population-Development-Environment in Namibia*. Edited by Fuller, B. and Prommer, I. University of Namibia, Multidisciplinary Research and Consultancy Centre.

Hosmer, D.W., Lemeshow, S., and Sturdivant, R.X. (2013). *Applied Logistic Regression.* John Wiley & Sons, Incorporated, New York.

Hossain, M. (2014). Examining the association between age at first sexual intercourse and HIV among women age 15-49 in Senegal. *Jaids-Journal of Acquired Immune Deficiency Syndromes, Volume: 67 Pages: 76-76 Supplement: 3 Meeting Abstract: P-A2 Published: NOV 2014*.

Johnson L.F., Dorrington R.E. and H. Moolla (2017). HIV epidemic drivers in South Africa: A model-based evaluation of factors accounting for inter-provincial differences in HIV prevalence and incidence trends. *Southern African Journal Of Hiv Medicine, Volume: 18 Issue: 1 Article Number: UNSP a695.*

Kalichman S.C., Ntseane D., Nthomang K. et al. (2007). Recent multiple sexual partners and HIV transmission risks among people living with HIV/AIDS in Botswana. *Sexually Transmitted Infections 2007;83:371-375.*

Kalipeni, E. and Zulu, L. (2008). Using GIS to model and forecast HIV/AIDS rates in Africa, 1986-2010. *Professional Geographer, vol. 60, no. 1, pp. 33-53.*

Kim A.A., Parekh B.S., Umuro M., Galgalo T., Bunnell R., Makokha E., Dobbs T., Murithi P., Muraguri N. and K.M. De Cock (2016). Identifying Risk Factors for Recent HIV Infection in Kenya Using a Recent Infection Testing Algorithm: Results from a Nationally Representative Population-Based Survey. *PLoS ONE 11(5): e0155498.*

Kulldorff, M. (1997). A spatial scan statistic. *Comm Stat Theor Meth. 1997;26:1481–1496.*

Lakew, Y., Benedict, S. and Haile, D. (2015). Social determinants of HIV infection, hotspot areas and subpopulation groups in Ethiopia: evidence from the National Demographic and Health Survey in 2011. *BMJ Open, 5(11), e008669.*

Larmarange J. (2013). Package ‘prevR’. <*https://cran.r-project.org/web/packages/prevR/prevR.pdf*>.Retrieved: 1.9.2015.

Larmarange, J., Vallo, R., Yaro, S., Msellati, P. and Méda, N. (2011). Methods for mapping regional trends of HIV prevalence from Demographic and Health Surveys (DHS), *Cybergeo: European Journal of Geography.*

Liu A, Cohen S, Follansbee S, Cohan D, Weber S, Sachdev D, et al. (2014) Early Experiences Implementing Pre-exposure Prophylaxis (PrEP) for HIV Prevention in San Francisco. *PLoS Med11(3): e1001613.*

Maffioli E.M. (2017). Is traditional male circumcision effective as an HIV prevention strategy? Evidence from Lesotho. *PLOS ONE Volume: 12 Issue: 5 Article Number: e0177076.*

Magadi, M. and Desta, M. (2011). A multilevel analysis of the determinants and cross-national variations of HIV seropositivity in sub-Saharan Africa: Evidence from the DHS. *Health & place, vol. 17, no. 5, pp. 1067-1083.*

Maher D., Waswa L., Karabarinde A. and K. Baisley (2011). Concurrent sexual partnerships and associated factors: a cross-sectional population-based survey in a rural community in Africa with a generalised HIV epidemic. *BMC PUBLIC HEALTH Volume: 11 Article Number: 651.*

Marston M., Harriss, K. and Slaymaker, E. (2008). Non-response bias in estimates of HIV prevalence due to the mobility of absentees in national population-based surveys: a study of nine national surveys. *Sexually Transmitted Infections 2008;84(Suppl I):i71-i77.*

Maughan-Brown, B., Kenyon, C. and Lurie, M.N. (2014). Partner Age Differences and Concurrency in South Africa: Implications for HIV-Infection Risk Among Young Women. *Aids and Behavior, vol. 18, no. 12, pp. 2469-2476.*

Messina, J.P., Emch, M., Muwonga, J., Mwandagalirwa, K., Edidi, S.B., Mama, N., Okenge, A. and Meshnick, S.R. (2010). Spatial and socio-behavioral patterns of HIV prevalence in the Democratic Republic of Congo. *Social Science & Medicine, vol. 71, pp. 1428-1435.*

Ministry of Health, Kenya (2005). *AIDS in Kenya*. Trends, Interventions and Impact, 7th edition, 2005 Kenya: National AIDS and STI Control Program, Ministry of Health, 2005.

Mishra, V., Assche, S.R., Greener, R., Vaessen, M., Hong, R., Ghys, P.D., Boerma, J.T., Van Assche, A., Khan, S. and Rutstein, S. (2007). HIV infection does not disproportionately affect the poorer in sub-Saharan Africa. *Aids, vol. 21, pp. S17-S28.*

Mmbaga, E. J. (2013). HIV Prevalence and Associated Risk Factors: Analysis of Change over Time in Mainland Tanzania. *DHS Working Papers 2013 No. 85. Demographic and Health Research.*

MOHSS (= Namibia Ministry of Health and Social Services) (2010). National Strategic Framework for HIV and AIDS Response in Namibia 2010/11 – 2015/16. Windhoek, Namibia. <*http://www.nationalplanningcycles.org/sites/default/files/country_docs/Namibia/namibia_2011-2016.pdf*>. Retrieved: 8.10.2017.

MoHSS (2014). Surveillance Report of the 2014 National HIV Sentinel Survey. Republic of Namibia, Ministry of Health and Social Services. Windhoek, Namibia. <*http://www.mhss.gov.na/files/downloads/12f_2014%20National%20HIV%20Sentinel%20Survey.pdf*>. Retrieved: 19.10.2017

MoHSS = The Namibia Ministry of Health and Social Services and ICF International (2014). The Namibia Demographic and Health Survey 2013, MoHSS and ICF International, Windhoek, Namibia and Rockville, Maryland, USA. <*https://dhsprogram.com/pubs/pdf/FR298/FR298.pdf*>. Retrieved: 8.10.2017.

Montana, L., Neuman, M. and Mishra, V. (2007). Spatial modeling of HIV prevalence in Kenya. *DHS Working Papers, no. 27.*

Moran P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika. 1950;14:17–33.*

Niragire F., Achia T.N.O., Lyambabaje A. and Ntaganira J. (2015). Bayesian mapping of HIV infection among women of reproductive age in Rwanda. *PLoS ONE. 2015;10(3):e0119944.*

Notkola V. and H. Siiskonen (2000). *Fertility, Mortality and Migration in SubSaharan Africa: The Case of Ovamboland in North Namibia, 1925-90.* London: Macmillan.

Notkola, V., Timaeus, I. M. and H. Siiskonen (2000). Mortality transition in the Ovamboland region of Namibia, 1930–1990. *Population Studies, 54 (2000), pp. 153–167.*

Notkola, V., Timaeus, I. M. and H. Siiskonen (2004). Impact on mortality of the AIDS epidemic in northern Namibia assessed using parish registers. *AIDS 2004, 18:1061–1065*.

NSA (2015). *Namibia 2011 Census: MIGRATION REPORT*. Namibia Statistics Agency.

NSA = Namibia Statistics Agency (2013): *Namibia 2011 Census Atlas*. Namibia Statistics Agency.

OpenStreetMap (2016). <*https://www.openstreetmap.org*>. Retrieved: 1.9.2015.

Preston, R. (1993). *The integration of returned exiles, former combatant and other war-affected Namibians, Final Report*. Namibian Institute for Social and Economic Research, University of Namibia, Windhoek.

Reniers G. and R. Tfaily (2012). Polygyny, Partnership Concurrency, and HIV Transmission in Sub-Saharan Africa. *DEMOGRAPHY Volume: 49 Issue: 3 Pages: 1075-1101.*

Sawers L. and A. Isaac (2017). Partnership duration, concurrency, and HIV in sub-Saharan Africa. *Ajar-African Journal Of Aids Research Volume: 16 Issue: 2 Pages: 155-164.*

Schaefer R., Gregson S., Takaruza A., Rhead R., Masoka T., Schur N., Anderson S.J. and C. Nyamukapa (2017). Spatial patterns of HIV prevalence and service use in East Zimbabwe: implications for future targeting of interventions. *J Int AIDS Soc. 2017 Feb 28;20(1):21409.*

Sepkowitz, K.A. (2001). AIDS – The First 20 Years. *The New England Journal of Medicine, Vol. 344, No. 23. June 7, 2001.* Massachusetts Medical Society.

Shelton, J., Cassell, M. and Adetunji, J. (2005). Is poverty or wealth at the root of HIV? *Lancet, vol. 366, no. 9491, pp. 1057-1058.*

Shemeikka, R. (1999). HIV/AIDS: The Major Demographic Challenge of Namibia. *Yearbook of Population Research in Finland 35 (1998-1999), pp. 65-81.*

Shemeikka, R. and V. Notkola (2005). Fertility Reduction Caused by HIV/AIDS and its Demographic Impact in Northern Namibia. In: *What are we doing there? Experiences and lessons learned from development cooperation in health care and social welfare (1990-2005)*. Publisher: National Research and Development Centre for Welfare and Health, pp.207-226

Shemeikka, R., Notkola, V., Kuhanen, J. and Siiskonen, H. (2007). HIV and Fertility Decline in North-Central Namibia 1980–2004. *Finnish Yearbook of Population Research 43 (2007–2008), pp.7–32.*

Siiskonen, H. (2007). *The Namibian mainline churches and HIV/AIDS in Aridity, Change and Conflict in Africa*: Proceedings of an International ACACIA Conference held at Königswinter, Germany October 1–3, 2003. Heinrich-Barth-Institut, Köln.

Siiskonen, H. (2009). HIV/AIDS eteläisessä Afrikassa: katastrofi yhteiskuntien kaikilla tasoilla. *Tieteessä tapahtuu, vol. 3, pp. 12-17.*

Simon, D. and Preston, R. (1993). Return to the promised land; the repatriation and resettlement of Namibian refugees, 1980-1990, in Black, R. And Robinson, V. (eds) (1993) *Geography and refugees; patterns and processes of change,* Belhaven, London.

Tanser, F., Barnighausen, T., Cooke, G.S. and Newell, M. (2009). Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *International journal of epidemiology, vol. 38, no. 4, pp. 1008-1016.*

Tanser, F., LeSueur, D., Solarsh, G. and Wilkinson, D. (2000). HIV heterogeneity and proximity of homestead to roads in rural South Africa: an exploration using a geographical information system. *Tropical Medicine & International Health, vol. 5, no. 1, pp. 40-46.*

Tenkorang E.Y. (2014). Marriage, widowhood, divorce and HIV risks among women in sub-Saharan Africa. *Int Health. 2014 Mar;6(1):46-53*.

The DHS Program (2017). <*https://dhsprogram.com/*>. Retrieved: 17.10.2017.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, Vol. 46, Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods (Jun., 1970), pp. 234-240.*

Turner, J.R. (2013). Area Under the Curve (AUC). In: *Encyclopedia of Behavioral Medicine* pp 125-126. Springer New York.

UNAIDS (2014). 2014 PROGRESS REPORT ON THE GLOBAL PLAN towards the elimination of new HIV infections among children by 2015 and keeping their mothers alive. <*http://www.unaids.org/sites/default/files/documents/JC2681_2014-Global-Plan-progress_en.pdf*>. Retrieved: 18.10.2017.

UNAIDS (2016). 2016 United Nations Political Declaration on HIV and AIDS: On the Fast-Track to Accelerate the Fight against HIV and to End the AIDS Epidemic by 2030. General Assembly Resolution 70/266. UNAIDS, Geneva. <*http://www.unaids.org/en/resources/documents/2016/2016-political-declaration-HIV-AIDS*>. Retrieved: 8.10.2017.

UNAIDS (2017a). AIDSinfo. UNAIDS. <*http://aidsinfo.unaids.org/*>. Retrieved: 8.10.2017.

UNAIDS (2017b). UNAIDS Data 2017. UNAIDS. <*http://www.unaids.org/sites/default/files/media_asset/20170720_Data_book_2017_en.pdf*>. Retrieved: 8.10.2017.

UNAIDS (2017c). Maximizing the potential of a new HIV prevention method: PrEP. <*http://www.unaids.org/en/resources/presscentre/featurestories/2016/november/20161101_PrEP*>. Retrieved: 22.10.2017.

UNAIDS = Joint United Nations Programme on HIV/AIDS (2000). *REPORT on the global HIV/AIDS epidemic*. UNAIDS, Geneva, Switzerland.

Wallace, Marion (2011). *A history of Namibia: from the beginning to 1990*. Columbia University Press, New York. 451 p. (Chichester, West Sussex)

Wand H. and G. Ramjee (2010). Targeting the hotspots: investigating spatial and demographic variations in HIV infection in small communities in South Africa. *J Int AIDS Soc. 2010;13(1):41.*

Webb, D. (1997). HIV and AIDS in Africa, Pluto Press, London.

Vissers D.C.J., Voeten H.A.C.M., Nagelkerke N.J.D., Habbema J.D.F. and de Vlas S.J. (2008). The Impact of Pre-Exposure Prophylaxis (PrEP) on HIV Epidemics in Africa and India: A Simulation Study. *PLoS ONE3(5): e2077.*

Voeten, H.A.C.M., Vissers, D.C.J., Gregson, S., Zaba, B., White, R.G., de Vlas, S.J. and Habbema, J.D.F. (2010). Strong Association Between In-Migration and HIV Prevalence in Urban Sub-Saharan Africa. *Sexually transmitted diseases, vol. 37, no. 4, pp. 240-243.*

World Bank (2009). Namibia – Country brief. The World Bank, Washington DC. <*https://openknowledge.worldbank.org/bitstream/handle/10986/2630/485960PUB0Na mi101Official0Use0Only1.pdf?sequence=1*> Retrieved: 25.3.2015.

World Bank (2015). Namibia Overview. <*http://www.worldbank.org/en/country/namibia/overview*>. Retrieved: 25.3.2015.

Zulu L.C., Kalipeni E. and E. Johannes (2014). Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994–2010. *BMC Infect Dis. 2014;14(1):285.*